NASA/TM-2016-219280

Data Mining Review of Team Benchmark Studies Related to Long-Duration Space Exploration Missions

Suzanne T. Bell, Ph.D.
Shanique G. Brown, M.A.
Tyree D. Mitchell, M. A.

DePaul University
College of Science and Health
1 E. Jackson, Chicago, IL 60604

National Aeronautics and
Space Administration

Johnson Space Center
Houston, Texas 77058

November 2016

NASA STI Program ... in Profile

Since its founding, NASA has been dedicated to the advancement of aeronautics and space science. The NASA scientific and technical information (STI) program plays a key part in helping NASA maintain this important role.

The NASA STI program operates under the auspices of the Agency Chief Information Officer. It collects, organizes, provides for archiving, and disseminates NASA's STI. The NASA STI program provides access to the NASA Aeronautics and Space Database and its public interface, the NASA Technical Report Server, thus providing one of the largest collections of aeronautical and space science STI in the world. Results are published in both non-NASA channels and by NASA in the NASA STI Report Series, which includes the following report types:

- TECHNICAL PUBLICATION. Reports of completed research or a major significant phase of research that present the results of NASA Programs and include extensive data or theoretical analysis. Includes compilations of significant scientific and technical data and information deemed to be of continuing reference value. NASA counterpart of peer-reviewed formal professional papers but has less stringent limitations on manuscript length and extent of graphic presentations.

- TECHNICAL MEMORANDUM. Scientific and technical findings that are preliminary or of specialized interest, e.g., quick release reports, working papers, and bibliographies that contain minimal annotation. Does not contain extensive analysis.

- CONTRACTOR REPORT. Scientific and technical findings by NASA-sponsored contractors and grantees.

- CONFERENCE PUBLICATION. Collected papers from scientific and technical conferences, symposia, seminars, or other meetings sponsored or co-sponsored by NASA.

- SPECIAL PUBLICATION. Scientific, technical, or historical information from NASA programs, projects, and missions, often concerned with subjects having substantial public interest.

- TECHNICAL TRANSLATION. English-language translations of foreign scientific and technical material pertinent to NASA's mission.

Specialized services also include creating custom thesauri, building customized databases, and organizing and publishing research results.

For more information about the NASA STI program, see the following:

- Access the NASA STI program home page at *http://www.sti.nasa.gov*

- E-mail your question via the Internet to help@sti.nasa.gov

- Fax your question to the NASA STI Help Desk at 443-757-5803

- Phone the NASA STI Help Desk at 443-757-5802

- Write to:
  NASA Center for AeroSpace Information
  7115 Standard Drive
  Hanover, MD 21076-1320

NASA/TM-2016-219280

Data Mining Review of Team Benchmark Studies Related to Long-Duration Space Exploration Missions

Suzanne T. Bell, Ph.D.
Shanique G. Brown, M.A.
Tyree D. Mitchell, M. A.

DePaul University
College of Science and Health
1 E. Jackson, Chicago, IL 60604

National Aeronautics and
Space Administration

Johnson Space Center
Houston, Texas 77058

November 2016

Acknowledgements

This work was supported by contract NNJ15HK18P, awarded to the first author from the National Aeronautics and Space Administration (NASA). All opinions expressed herein are strictly those of the authors and not necessarily those of NASA. We especially thank Alan H. Feiveson and Lauren B. Landon for providing insightful comments on an interim report, Alla Vinokhodova and Gloria Leon for helping us obtain some limited circulation publications with valuable data, and Kim Binsted and Pete Roma for sharing unpublished data. Question or comments about the report should be directed to Suzanne T. Bell, Department of Psychology, DePaul University, 2219 N Kenmore Ave, Chicago, IL 60614, USA; (t) 773.325.4246, SBELL11@depaul.edu

Available from:

NASA Center for AeroSpace Information
7115 Standard Drive
Hanover, MD 21076-1320

National Technical Information Service
5301 Shawnee Road
Alexandria, VA 22312

Available in electronic form at http://ston.jsc.nasa.gov/collections/TRS

ABSTRACT

As part of its integrated research plan, NASA seeks to manage the risk of performance and behavioral health decrements that result from inadequate cooperation, coordination, communication, and psychological adaptation within a team. As a part of this risk, NASA has identified a specific gap in understanding the key threats, indicators, and life cycle of teams that will embark on autonomous, long-duration and/or long-distance space exploration missions (LDSEM), Team Gap 1. An increasing number of studies have examined teams in LDSEM-analog environments (e.g., chamber simulations, Antarctic winter-overs) and can help inform this gap. We conducted a quantitative review of existing studies benchmarking team factors in LDSEM-analog environments to summarize the existing quantitative evidence on team functioning in such environments and to identify gaps to address in future research. This is the final report for our project (NASA contract NNJ15HK18P). We describe the methodology we used, the results of our efforts, and important research priorities related to Team Gap 1 moving forward.

EXECUTIVE SUMMARY

We sought to conduct a quantitative review of team benchmark studies conducted in long-distance space exploration mission (LDSEM) analog environments. We included research conducted on teams that live and work together, and we coded specific variables such as isolation and confinement to indicate fidelity to the LDSEM context. Through extensive search efforts, we were able to obtain a total of 70 sources that provided quantitative data on team functioning in LDSEM-analog environments. These sources informed two data sets: an effect size data set and a benchmarking data set. Eleven sources (e.g., journal articles, technical reports) provided enough data (team $n \geq 5$) to generate 253 team-level effect sizes for team-factor (e.g., team cohesion) and team-outcome (e.g., team performance) relationships. Fifty-nine sources included team functioning data from fewer than five teams; from these sources, we were able to glean 1,051 data instances (data collected on one or more variable at a particular time point) to benchmark team functioning in LDSEM-analog environments over time. The data were not appropriate for traditional meta-analytic approaches; accordingly, we explored the data sets using two other techniques.

First, we calculated a weighted average of the effect size from the LDSEM-analog environment (i.e., the local validity estimate) and a relevant meta-analytic estimate from the extant literature where available, as a minimum-variance estimate. We used this approach to generate the best possible estimation of the predictor and outcome relationship in the context, given the data, and to report credible intervals of the estimates. Results suggest that the team cohesion and team performance relationship may be operating differently in isolated and confined environments (ICE; e.g., Antarctic stations, lab research where participants were constrained to an ICE), and that, given the available data, we can say very little about the magnitude and direction of the effect. Conversely, similar to team cohesion and team performance effects observed in the extant literature on traditional teams (see Beal, Cohen, Burke, & McLendon, 2003), our results suggest, with 95% certainty, that the true team cohesion and team performance relationship in the local population of the non-ICE studies (e.g., firehouses, special operations teams) examined in our review is positive, and small to large. Although limited, these results underscore the importance of specifically studying how team cohesion operates in isolated and confined spaces. Information on other

relationships (e.g., age homogeneity and team performance; education level homogeneity and team performance) is limited and reflects a similar inability to estimate the true population validity of team factors on team functioning in ICE, given the existing data.

Second, we explored our team benchmarking data set by generating figures that depict team functioning over time. We were able to generate figures of comparable data from multiple analogs for team efficiency, team conflict, team communication, and team mood over time. These data cannot speak to cause and effect; however, they may be useful for benchmarking typical or atypical team functioning. The data revealed interesting trends. For example,

- Consistent team efficiency over time was most common; decreased team efficiency over time was atypical.
- The number of intra-team conflicts varied over time; however, by 40% of the mission-completion point (in our data set, equivalent to at least 90 days of a mission), all teams reported at least one conflict within the team or with mission control.
- Trends for commanders' written communication with mission control were consistent with the "psychological closing" phenomenon described by Gushin and colleagues (2012) in that commanders' reports decreased in length over time.
- While there was not consistent support for the third-quarter phenomenon for team mood operationalized as total mood disturbance or positive affectivity, three of seven LDSEM-analog teams reported increased negative affectivity in the third quarter. Two teams (i.e., Mars 500 and an Antarctic team) reported total mood disturbance data for a year or longer; both of these teams reported a spike in team total mood disturbance about one year into the mission.

In the final section of the report, we suggest an agenda for future research conducted on team functioning in analog environments; tables summarize the research recommendations in priority order. We summarize the recommendations here:

- Longitudinal research (i.e., durations longer than 90 days and ideally longer than 1 year) in environments with high fidelity to LDSEM to study the effects of isolation and confinement on team functioning should be conducted. Experimental and quasi-experimental research conducted in LDSEM-analog environments that can identify triggers of changes in team functioning is also needed.
- Given that high fidelity environments to LDSEM, particularly ICE, have a small team-level sample size, it is **essential** that data are collected in such a way that they can be aggregated and compared across multiple simulations.
    - Data on key team constructs (e.g., team efficiency, communication, mood, and cohesion) should be collected with a common set of measures (i.e., a "common core"). At a minimum, the team common core should include measures that have been and continue to be systematically collected in important, longer-duration simulations (e.g., MARS 500, Simulated Flight of International Crew on Space Station [SFINCSS]) conducted by Institute of Biomedical Problems [IBMP], Russian Academy of Sciences, so that new data are directly comparable across simulations and with previously collected data.
    - Clear data sharing guidelines should be established. It is essential that NASA has access to data collected as part of the common core from all NASA-sponsored, LDSEM-analog data-collection efforts within a reasonable time frame, likely shorter than the typical journal publication timeframe. Specific innovations developed by individual researchers could be governed by a separate set of, and in

some cases, more restricted data-sharing guidelines than would govern the common core.
    - Innovation can be encouraged alongside, or in addition to, the common core. When the innovations pertain to the measurement of a construct included in the common core, data should be collected to determine the comparability of the new and old measures to allow for the comparison of data in aggregate. A scale homogenization process may provide a possible path for determining equivalence.
- Although progress is being made, there is still a significant gap in understanding the key threats, indicators, and life cycle of a team for LDSEM. These areas are in need of continued research so that we can accumulate knowledge across different teams with varying compositions, under different conditions. In addition to research currently underway (e.g., team composition, team cohesion, communication), we suggest some areas in need of more significant exploration.
    - Research on affect in LDSEM-analog environments exists; however, it is generally focused at individual team member's affect. Applying a team lens to research on team affect (e.g., the extent to which and what factors amplify emotional contagion) and team affect management will allow for a more complete understanding of the role of affect in LDSEM team performance and crew member well-being.
    - We thought the decline in team efficiency during the Human Behaviour in Extended Spaceflights (HUBES) simulation, as well as the dip in team efficiency for one of the SFINCSS teams was interesting. Researchers noted the lack of an established leadership structure and the conflict between team members as possible explanations for the performance effects (e.g., Sandal, 2001; Vinokhodova et al., 2001). Given the possibility of shared leadership for LDSEM, an exploration of the conditions necessary for teams to establish a workable leadership structure, triggers of status conflict, and approaches to resolving status conflict in multicultural crews is an important avenue of research.
    - A critical component to composing, training, and providing countermeasure support to crews must include a consideration of team resilience. Future research should consistently measure team resilience in response to specific manipulations of acute stressors (e.g., disagreement over the inclusion of mission control to resolve a specific conflict), as well as resilience in response to the subtle changes (e.g., change in team mood) that will occur during a team's life cycle.

Keywords: human behavior; human performance; psychological effects; psychology; teams.

TABLE OF CONTENTS

Abstract iii

Executive Summary iii

Background 1

Method 2

Inclusion Criteria 2

Literature Search 3

Description of Variables 4

Coder Training and Agreement 5

Data Sets 5

Analytical Strategy 7

Results 10

Weighted Averages 10

Figures Benchmarking Team Functioning Over Time 13

Gaps and Research Recommendation 26

Collecting Data in ICE environments Such as the Antarctic and Chamber Simulations Such as HERA 26

Sponsoring Research that Allows for the Culmination of Data 28

Gaps and Research Recommendations Regarding Specific Team Factors 35

References 41

Appendix A – List of Search Terms 50

Appendix B – List of Specific Journals Searched 51

Appendix C – List of Team Factors & Included Studies 53

Appendix D – Why the Current Data Set Does Not Lend Itself to Traditional Forms of Meta-Analysis 72

Appendix E: Acronyms 73

BACKGROUND

Given the challenges associated with future long-distance space exploration missions (LDSEM), NASA seeks to minimize risks to human health and performance, including the risk of team performance decrements that may result from inadequate cooperation, coordination, communication, and psychosocial adaptation. Much is known about the functioning of teams (see Mathieu, Maynard, Rapp, & Gilson, 2008, for a review); however, most of this research is from teams that operate in traditional workplaces (e.g., office settings, production plants). Less is known about how teams function in extreme environments, such as the environments that LDSEM crews would encounter. NASA identifies the gap in our collective knowledge regarding the key threats, indicators, and life cycles for LDSEM crews as Team Gap 1. We address this gap in our research.

LDSEM crews will be expected to operate in an extreme environment. The current design reference manual (DRM) suggests that LDSEM crews are anticipated to live and work together in isolation and confinement for up to 30 months. LDSEM crews are also expected to experience up to 22 minutes of communication delay (one-way), limited video communication, and high autonomy, which may increase over time. These unique challenges are likely to impinge on team functioning. Because of this, compared to research on teams in traditional settings, research conducted on teams facing similar circumstances expected for LDSEM (e.g., LDSEM analogs) is thought to provide better insight into LDSEM team functioning. Research on teams that operate in extreme environments, such as Antarctic, Arctic, and high-fidelity space simulations (e.g., HERA, MARS-500, share many of the same characteristics expected for LDSEM (e.g., isolation, confinement) and are often used as LDSEM analogs to help identify important issues for long-duration spaceflight (Palinkas, Gunderson, Johnson, & Holland, 2000).

An increasing number of studies have examined teams in LDSEM-analog environments. A data-mining effort and review of existing quantitative studies benchmarking team factors in LDSEM-analog environments is necessary to summarize the existing quantitative evidence on team functioning and to identify gaps to address in future research. The primary purpose of our research was to provide an overall picture of the available data on team functioning and performance in LDSEM-analog environments. We answer three primary questions in our data-mining review of team research conducted in LDSEM-analog environments.

(1) What factors related to team functioning have been studied in analog environments?

(2) What do these studies reveal about team functioning in LDSEM-analog environments?

(3) What do we recommend as research priorities for NASA moving forward?

This report summarizes the results of our research and is organized as follows. First, we describe the search strategy we used to locate research related to team functioning in spaceflight and analog environments, our inclusion criteria, and our coding strategy. Second, we summarize the available evidence on team functioning from spaceflight and LDSEM-analog environments. We do so by providing weighted averages of the local validities and meta-analytic estimates drawn from the existing literature as minimum-variance estimates, and figures depicting team functioning over time. Finally, we identify critical gaps in the research and provide research recommendations.

METHOD

INCLUSION CRITERIA

We sought to be as inclusive as possible in our data-mining review while also striving to ensure that the data we reviewed were relevant to understanding team functioning in an LDSEM environment. Three general issues should be noted. First, our review is focused on team-level (as opposed to individual-level) functioning, and accordingly, we identified and included only team-level data. Teams are defined as two or more individuals who share a common goal and perform interdependent tasks that are relevant to an organization (Kozlowski & Bell, 2013). This team-level focus is important both empirically and theoretically. Empirically, levels of analysis (i.e., at the individual-level, at the team-level) can greatly influence sample size. For example, a study of 10 teams with 6 individuals on each team would have a team-level sample size of 10 and an individual-level sample size of 60. Combining levels of analysis when using techniques such as meta-analysis is problematic given the dependence on sample size for computing important formula components such as standard error or variability. Individual-level effects would be disproportionately weighted in the estimates. Theoretically, relationships observed at one level cannot be assumed to generalize to a different level (e.g., the team-level). This is called a cross-level fallacy (Klein, Dansereau, & Hall, 1994; Kozlowski & Klein, 2000; Rousseau, 1985). Thus, we included data on team functioning at the team-level. Individual-level functioning (e.g., individual-level adaptation) in LDSEM-analog environments is important but is outside the scope of this report. Therefore, we excluded articles that reported individual-level data that were not tied to a particular team (e.g., Bartone, Johnsen, Eid, Brun, & Laberg, 2002), or that were tied to a large polar station (>40 people) but not to a team or a small station (e.g., Doll & Gunderson, 1971; Palinkas, Gunderson, & Burr, 1989).

Second, our focus was on research conducted on teams in LDSEM-analog environments. As mentioned, research on teams that operate in extreme environments, such as Antarctic, Arctic, and high-fidelity space simulations (e.g., HERA, MARS-500), share many of the same characteristics expected for LDSEM (e.g., isolation, confinement) and are often used as LDSEM analogs. Defining an LDSEM-analog environment has challenges, however, because a particular extreme environment (e.g., Antarctic winter-overs) may only share some of the same characteristics expected of LDSEM. For example, one context may have participants who live and work in isolated and dangerous environments but also have some crew rotation. Alternatively, participants may live and work together for a long duration and complete their work in a hostile environment but are not isolated and confined (e.g., firefighters). For the purposes of this review, LDSEM-analog research was defined as research in which members of the focal team (e.g., the "crew" analog) live and work together for a period of time. We included military teams when they were expressly described as intact teams (e.g., combat teams such as Lim & Klein, 2006; Ko, 2005) even if the research did not explicitly mention that the unit lived together. We did not include military or firefighter training exercises when it was unclear whether the team lived together either while at training or while not at training (e.g., Hirschfeld & Bernerth, 2008; Oser, McCallum, Salas, & Morgan, 1989). We also excluded sources that included data on children (e.g., Tyerman & Spencer, 1983). We chose a broad definition of LDSEM-analog environments and then coded the specific features that would allow for fidelity comparisons between the study and the anticipated LDSEM environments (e.g., isolation and confinement, workload, and working in a dangerous environment) as moderators.

Third, we included sources that reported quantitative data (i.e., we excluded descriptive case studies and narrative reviews). Typically, meta-analysis is preferred for integrating estimates

of the same relationship of interest across studies; it allows us to generate cumulative knowledge about a set of studies. The benefits of meta-analysis over narrative reviews have been widely noted (see Glass, 1976; Arthur, Bennett & Huffcutt, 2001). Early in our review process, however, we suspected that most studies conducted on teams in analog environments would not lend themselves to traditional meta-analysis; frequentist meta-analytic techniques can be inappropriate when a limited number of studies have examined a particular relationship or when sample sizes or data do not permit the calculation of an effect size, for example, when data are only reported for a single team. Further, a review of the analog research at the individual-level determined that meta-analytic techniques were inappropriate (e.g., Shea, Slack, & Keeton, Palinkas, Leveton, 2011). Because we suspected the same may be true at the team-level, we decided to retain sources whether or not we could compute an effect size, provided that there were quantitative data reported. When a source reported data on 5 or more teams, we coded an effect size, either r or Spearman's ρ, for the relationship between a predictor and an outcome related to team functioning. For sources with a team sample size less than 5, we coded quantitative data such as means (or another team-level representation) and within-team standard deviation, when available, for team functioning variables across time. We then reviewed all coded studies to determine the most appropriate approach for benchmarking team factors so as to provide an overall picture of the team functioning and performance, given the available information. In the next section we describe our literature search, inclusion criteria, and coding process in more detail.

LITERATURE SEARCH

We used four search strategies to obtain quantitative research on teams in LDSEM-analog environments: (1) a thorough literature search using multiple databases; (2) a search of the NASA taskbook to identify potential data and technical report leads and researchers to contact; (3) posts to listservs and requests to individual researchers thought to have collected relevant data in analog environments; and (4) a review of reference lists of key articles and other reviews of similar domains (e.g., Schmidt, 2015). The search process began in January 2015 and included all dates to the present (search date). Researchers were contacted in May 2015. The strategies are described in this section.

We conducted a targeted search of 13 specific databases that were selected based on their success in identifying articles for a previous LDSEM-analog review we completed. The databases included space agency databases such as the NASA, ESA, and JAXA technical report repositories, Military and Government Collection, and more general databases such as Google Scholar and EBSCOhost. We included a broad range of search teams, including team functioning keywords (e.g., performance, cohesion) and environment or simulation characteristics (e.g., Antarctic, Mars 500), to identify articles that examined team functioning in analog environments. See Appendix A for a detailed list of our search terms and databases searched. We also searched specific journals; see Appendix B for a list of the specific journals searched. In addition, we reviewed articles from a list of LDSEM-analog studies identified previously by NASA's Behavioral Health and Performance (BHP) unit. We also reviewed the reference lists of recently completed reviews of similar domain areas (e.g., Bell, Brown, Abben, & Outland, 2015; Palinkas et al., 2011; Schmidt, 2015) and the reference lists of recent technical reports on team research funded by NASA (Bell, Brown, Outland, & Abben, 2015; Burke & Feitosa, 2015; DeChurch & Mesmer-Magnus, 2015; Gibson et al., 2015; Morgeson, 2015; Smith-Jentsch, 2015). Finally, we reviewed the reference lists of articles from which we were able to generate an effect size (e.g., Emurian & Brady, 1984; Gunderson & Ryman, 1967).

We also sought unpublished or in-press research on team functioning conducted in analog environments. To do this, Suzanne Bell e-mailed 29 researchers thought by us or by BHP to conduct

research in the domain of interest (e.g., Kanas, Vinokhodova, Sandal, Leon, Kanki, Roma, Bishop, Binsted). We also requested research via posts on listservs including those sponsored by Science of Team Science, INGRoup, Research Methods, Organizational Behavior, and the human resources divisions of Academy of Management. These efforts resulted in 5 additional books, articles, and datasets sent by researchers. In total, we identified approximately 309 sources (e.g., books, technical reports, dissertations, journal articles, and conference papers) for possible inclusion.

Next, we conducted a more detailed review of the identified sources and applied the inclusion criteria previously mentioned. To better understand the nature of the available data, we sorted the 309 sources into three categories: (1) sources that included quantitative data with a team-level sample size of 5 or greater, for which a team-level effect size between a predictor and criteria related to team functioning could be generated; (2) sources that included quantitative data on fewer than 5 teams or only data for one variable over time; and (3) sources that did not provide relevant data for our quantitative review. Sources in the third category were excluded from further review. The decision to exclude an article was agreed upon by at least two members of the research team. Applying the specified inclusion criteria resulted in 11 different sources (journal articles, technical reports) that provided enough information to calculate effect sizes representing the relationship between a predictor and a criterion related to team functioning, as well as 59 different sources that reported quantitative data on team functioning over time in LDSEM-analog environments but that did not enough data to calculate an effect size. We coded the 70 sources for fidelity characteristics and other moderators along with the quantitative data on team functioning. In the next section, we provide details about the variables coded and information about the coding process, including information about coder training.

DESCRIPTION OF VARIABLES

Two coding forms were created: one for coding effect sizes and one for coding data (e.g., means and standard deviations) related to team functioning over time. Coding forms were similar in that both captured characteristics of the source, the sample, fidelity characteristics, and information about the predictor and/or criteria. In addition, a codebook with definitions of the variables and descriptions for the different categories for each variable was developed.

Study design was coded as: (a) descriptive, (b) correlational, (c) quasi-experimental, and (d) experimental. We coded the degree of similarity between the sources' samples and the anticipated characteristics of LDSEM crews in terms of gender, national background, and education. We also coded the fidelity of the sources' characteristics to the characteristics expected in LDSEM environments. Examples of the characteristics we coded include mission duration, crew size, workload amount, workload variability, whether the crew participated in a multi-team system (e.g., interacted with mission control [MC]), how the crew communicated with MC and whether there was a time delay, level of crew autonomy, whether the team was static or experienced a change in membership during the time of the mission, and whether the teams performed work in a dangerous environment during the period of data collection. Studies were coded as occurring in dangerous environments when the setting had features that required individuals to use special equipment (e.g., winter-overs in Antarctic) or posed an imminent threat (e.g., polar bear threat). Studies that collected data during training exercises were coded as not dangerous unless the authors explicitly mentioned a dangerous training environment. Studies were coded as isolated when team members were limited in physical interaction with outside parties for a substantial period of time during the study, and confined when they primarily operated in a highly restricted space. For example, winter-overs in small Antarctic stations or space simulations were coded as an isolated and confined environment. Autonomy was coded as high, moderate, low, or not reported. Many

studies did not describe the level of autonomy in great detail and were coded as 'not reported'. Mission length was coded as the total of number of days in the team's life cycle. Ongoing teams such as firehouses (e.g., Kniffin, Wansink, Devine, & Sobal, 2015) were coded to the max of the distribution (e.g., 730 days).

We coded fidelity characteristics when they were described or could be reasonably assumed by two independent coders, given the descriptions provided in the sources. We used the Internet to locate factual information about specific simulations or Antarctic stations to complete missing fidelity information where possible. We gleaned an effect size estimate (either a Spearman's ρ or a Pearson's correlation) between team functioning factors (e.g., conflict) and outcomes (e.g., performance) when available; if an effect was not given, we calculated one if enough information was presented. When there was an insufficient sample size or only one variable reported, we coded the reported team-level effect (e.g., means and standard deviation) over time. We included data that were presented numerically as well as those presented in figures, except when the approximate value reported in the figure could not be reasonably estimated (e.g., due to ambiguities in labeling of the axis).

CODER TRAINING AND AGREEMENT

Shanique Brown and Tyree Mitchell, both post-master's industrial and organizational psychology doctoral students who had an advanced understanding of team functioning and who became familiar with the LDSEM context (e.g., by attending IWS, reading articles on context, reviewing DRM) served as coders for this study. Suzanne Bell trained the coders on the coding scheme described in the previous section. Coders first received a coding sheet and a codebook that provided descriptive information about each category of variables. All three authors then used the codebook and coding sheet to independently code three articles. The three authors met to discuss the coding, observe areas of agreement and disagreement, and make modifications to the coding sheet and codebook. Next, all three authors recoded the initial set of articles to help establish a frame of reference that incorporated the modifications made to the coding documents. Disagreements about the coding were resolved during a follow-up meeting using a consensus approach. After the second round of coding, a common set of 5 articles was coded to determine the efficacy of the coding process and to establish decision rules. When there was little disagreement (i.e., less than 3 disagreements across the variables coded in the studies), two coders coded the remaining articles. A randomly sampled common set of coded articles indicated that initial agreement, prior to the consensus meeting between coders, was relatively high (mean agreement of 87% on the variables that were coded). Discrepancies between the two coders were discussed and agreement was reached using a consensus approach. When consensus could not be reached with certainty between the two coders, the coders met with Suzanne Bell to discuss how the characteristic in question should be coded. After the coding was completed, we inspected the data sets to better understand the nature of the data, to determine the appropriateness of meta-analysis for summarizing the data, and to determine the best way to summarize the available evidence.

DATA SETS

Eleven sources (e.g., journal articles, technical reports) provided enough data (team $n \geq 5$) to generate 253 team-level effect sizes that represent a team factor (e.g., team cohesion) and team outcome (e.g., team performance) relationship. Fifty-nine sources included data on team functioning from fewer than 5 teams; from these sources we were able to glean 1,051 data instances (i.e., data collected on one or more variable at a particular time point) to benchmark team functioning in LDSEM-analog environments over time. These two data sets provide information on which factors related to team functioning have been studied in analog environments. Appendix C, provided as an

attachment, contains fidelity information (Table 1) on all sources, a list of team effect sizes by study (Table 2), and a list of team factors for studies providing benchmark data (Table 3).

Fidelity Information

Table 1 of Appendix C includes fidelity information for all sources coded during our review. Studies are listed in the tables based on study settings (e.g. natural analogs, simulations, lab). We also provide information for each study, including study date, crew size, sample demographics, whether the crew members were reported as being familiar with each other or not before the time of the study, whether the environment the crew operated in was dangerous, and the reported mission length. We describe the characteristics of the studies included in our review in more detail in the remainder of this section. However, incomplete information provided in some studies prevented us from calculating fidelity scores that were meaningful (as done in Bell et al., 2015; Palinkas et al., 2011).

Of the 11 studies that contributed to our effect size data set, 4 studies included data from natural analogs (e.g., Antarctic), 6 were from military or firehouse teams that met our inclusion criteria, and 1 was a lab study in which participants were isolated and confined to a programmed environment for 10 days. Half of the studies were published prior to 1985, including all 5 studies in which teams were in isolation and confinement. The average crew size for these studies ranged from 4 to 18 members; there was one exception in which a study included a 40-member station (Gunderson & Nelson, 1963). With an expected crew size of 6 for long-duration missions (DRM), the size of crews included in the effect size data set generated from ICE is generally larger than would be ideal for comparison. The characteristics of crew members are another concern regarding study fidelity. Crews for LDSEM are expected to be diverse in terms of gender and cultural and professional backgrounds (DRM). For studies that contributed an effect size, team members were, on average, in their late 20s and mostly male. There was some professional diversity reported in Antarctic studies, but most studies were more homogenous in terms of professional background (e.g., military and firefighting teams) or did not report that information. The DRM also reports that crew members for LDSEM may have 2.5 to more than 3.5 years of familiarity with each other through training. For the most part, studies included in the effect size data set did not report the familiarity of team members prior to the formation of the team. Regarding danger, almost half of the studies included teams dealing with a dangerous environment, which is consistent with conditions expected of LDSEM. Finally, mission duration across the studies ranged from 10 to 730 days, which is less than the 900 days expected for long-duration missions (DRM).

For the benchmarking data set, 59 studies reported quantitative data on team functioning, including data from 17 different space simulations (e.g., Mars 105, Experimental Campaign for the European Manned Space Infrastructure [EXEMSI]. Simulated Flight of International Crew on Space Station [SFINCSS], some with multiple teams), 20 different natural analog teams (e.g., Antarctic expeditions, Scott base, some reporting multiple winter-overs), 4 lab studies, and 2 ISS spaceflight teams. Studies included in our benchmark analyses had more variability in fidelity characteristics compared to the effect size data set. Most of these studies occurred after the year 2000 (70%). Team size ranged from 2 to 28 members, and approximately 75% of the studies had a crew size of 6 or fewer. Where reported, the average age of crewmembers ranged from 27 to 36 years. Regarding gender, almost 75% of the studies reporting the gender composition indicated that teams were either all or mostly men. Approximately 50% of the benchmark studies report some professional diversity, and almost 40% reported diversity on national background. Several articles did not report the familiarity of crewmembers prior to the mission; however, around 45%

of the articles mentioned some amount of familiarity among team members. A third of the studies operated in a dangerous environment (approximately 30%) and mission length ranged from 7 to 520 days.

Effect Size Data Set

Tables 2 and 3 in Appendix C summarize which team factors have been studied in analog environments. Table 2 includes the predictor and criterion relationships, along with their effect sizes, related to team functioning across 11 articles. We refer to this as our effect size data set. Of the 253 effects across 11 different sources, the majority of effects (i.e., 102 effects across 9 studies) represented the relationship between a predictor and team performance. Forty-seven effects across 6 studies represented the relationship between a predictor and cohesion or compatibility, and the remaining effects represented a variety of outcomes that differed across studies. The specific predictor and criterion relationship examined varied across studies. Predictors included inputs, emergent states, and team process variables, such as personality (e.g., Gunderson & Ryman, 1967), values (e.g., Ko, 2005), compatibility, cohesion (e.g., Gunderson & Nelson, 1963), mental models (e.g., Lim & Klein, 2006), leadership (e.g., Lim & Ployhart, 2004), exploratory search, and planning (e.g., Knight, 2015). Outcome variables included performance effects (e.g., accomplishment, accuracy, time to completion), emergent states, team processes, and other team functioning such as cohesion, team mood, egalitarian atmosphere, viability, and cooperation. The data were largely dependent (i.e., the 253 effects came from only 11 different sources), and a variety of predictor and outcome relationships were examined. Only the relationship between measures of cohesion (e.g., compatibility, spending time together) and team performance was examined in more than 3 independent samples ($k = 6$).

Benchmarking Data Set

Table 3 of Appendix C includes sources that provided quantitative data on team functioning, but did not provide data that could be used to generate a team-level effect size. We refer to this as our benchmarking data set. Team factors included emergent states, team processes, outcomes, and additional team functioning markers. For example, emergent states included team cohesion (e.g., Allison, Duda, & Beuter, 1991; Vinokhodova, Gushchin, Eskov, & Khananashvili, 2012), and team processes included conflict and tension (e.g., Leon, List, & Magor, 2004; Wood, Lugg, Hysong, & Harm, 1999). Outcomes included performance (e.g., Emurian & Brady, 1984) and more subjective outcomes such as satisfaction (e.g., Bhargava, Mukerji, & Sachdeva, 2000; Leon et al., 2004). Finally, other functioning markers, such as team mood (e.g., Bishop, Kobrick, Battler, & Binsted, 2010; Kahn & Leon, 2000; Steel, 2001), were commonly reported in analog studies.

ANALYTICAL STRATEGY

Although we were able to locate quite a bit of data for our review, the data did not lend themselves to traditional meta-analytic techniques. Please see Appendix D for a more detailed discussion of why meta-analysis was deemed to be inappropriate for the effect size data set. Given the inability to broadly apply traditional meta-analytic techniques to the data, we used two other approaches for summarizing the two data sets. First, when the team factor and team outcome relationship could be represented using an effect size, we calculated a weighted average of the effect size and the relevant meta-analytic estimate in the extant team literature as a minimum-variance estimate. We use this approach to quantify our expectations of team functioning in the LDSEM-analog environment given the available data and our uncertainty surrounding those expectations. Second, when the number of teams included in the study was too few to generate an effect size, and when data across studies were comparable, we descriptively summarized the data on team functioning over time

via a series of figures. In the remainder of the analytic strategy section, we describe our weighted averages approach. We follow this with a section containing the weighted averages results, as well as the series of figures generated from the benchmarking data set.

We sought to provide the best estimate of team predictor-team criterion relationships in the LDSEM-analog environment, given the available data, and to quantify the uncertainty around the estimates. To do so, we generated improved estimates of the true population relationships in an LDSEM-analog environment by using an inverse weighting of the variances of the validity coefficients from the LDSEM-analog studies and the meta-analytic effects of the same predictor-outcome relationships in the extant literature. We also calculated the average inaccuracy of the estimates and used these to create 95% credible intervals to quantify the uncertainty of the estimates. We used this approach as a means of balancing the precision that meta-analysis can provide in estimating a relationship across multiple settings with the high uncertainty (especially due to small sample sizes, etc.) but localness that a specific effect size generated in an LDSEM-analog environment can provide. We explain this further in the next paragraph.

A broad and extensive literature base exists on team functioning (see Mathieu et al., 2008 for a review). Numerous meta-analyses summarize the extant literature on team functioning and quantify relationships such as the relationships between shared cognition and team performance (DeChurch & Mesmer-Magnus, 2010), team composition and team performance (Bell, 2007), and team cohesion and team performance (Beal et al., 2003). Much of the data included in these meta-analyses are from teams in more traditional settings, such as office project teams and military action teams. Given the extreme circumstances that LDSEM crews are expected to face, LDSEM-analog environments are thought to provide a better sense of team functioning for LDSE, and in that sense, serve as "local" validity studies conducted with the intended population. Local validity studies in general, but especially in the case of analog research, tend to have limited sample sizes. Because of this, there is uncertainty regarding how well an observed effect in a given sample reflects the true population correlation. Our weighted averages approach combines our prior knowledge (e.g., meta-analytic effects on team functioning conducted in traditional teams) with new knowledge (e.g., a local validity study in an analog environment) to arrive at a revised estimate of the unknown quantity (i.e., the true population validity of a team factor and outcome in the analog environment). It also provides an estimate of uncertainty surrounding that new estimate.

Our approach is equivalent to the empirical Bayesian approach described in Newman, Jacobs, and Bartram (2007). For reasons discussed in this paragraph, however, we refer to our approach as weighted averages. First, we acknowledge that some of our LDSEM-analog validity estimates, particularly those from isolated and confined environments are based on small sample sizes, which may influence the normality of the local validity distribution. Newman et al.'s (2007) local validity Bayesian estimation approach is only regarded as Bayesian when the distribution of the local estimate is normally distributed. Because it is not possible for us to test this assumption without access to raw data, we refer to our approach as taking a weighted average.1 Second, we acknowledge that it was not possible for us to estimate potential bias due to certain moderators such as whether the analog study was conducted in an ICE or non-ICE environment. However, Newman et al. (2007) indicates that the accuracy of their local validity Bayesian estimation approach holds true even in the presence of true moderators (e.g., teams that perform in ICE environments, for example, where

[1] We greatly appreciate Dr. Alan H. Feiveson who drew our attention to potential issues with bias and assumptions of normality and who informally ran some simulations which suggested that normality and variance assumptions for the z-transformation of correlations were decent for $n = 7$, which is the smallest sample size on which our effect size data were based, *if the original data were distributed as bivariate normal*.

the ICE/non-ICE context moderates the observed predictor and outcome relationship). Even so, we acknowledge that because we cannot assess or model the bias that may be present due to combining a local effect size from an ICE environment with a meta-analytic effect from non-ICE environments, we are trading an unknown amount of bias to generate a minimum variance estimate. If raw data were available, it would be better to do a full Bayesian analysis that takes into account sampling variability at the local level, as well as any bias in using a meta-analytic estimate based on the broader team literature as the prior distribution. Given the limitations of available data, however, we believe our weighted averages approach provides the best estimate of the team predictor and outcome relationships in the specific LDSEM-analog environment.

We relied on equations 1, 2, 3, and A12 from Newman et al. (2007) in forming our weighted averages. We used estimates from meta-analyses in the extant literature (e.g., Beal et al., 2003; Bell, Villado, Lukasik, Belau, & Briggs, 2011: LePine, Piccolo, Jackson, Mathieu, & Saul, 2008) to inform the prior probability distribution. We only generated an estimated distribution of the true population local validity when there was a relevant meta-analytic effect reported in the extant literature that could inform our prior distribution. This limited the number of relationships we estimated and narrowed the effects to team performance as the outcome. Further, even with performance as the outcome, there were a number of relationships for which we could not locate relevant meta-analyses; the relationships between LMX and team performance and between personality characteristics (e.g., conscientiousness) of the leader and team performance are examples. We also did not locate relevant meta-analyses for many of the personality and needs variables examined by Gunderson and Ryman (1967), such as wanted affection and nurturance personality. Finally, there were two estimates from military teams (e.g., shared mental models from Lim and Klein [2006] and collectivism from Ko [2005]) that were already included in meta-analysis that would have been used in the calculation of the weighted averages (DeChurch & Mesmer-Magnus, [2010]; and Bell [2007] respectively). Thus, we did not estimate local validity of these two estimates using our weighted averages approach. Appendix C contains all effect sizes observed in primary studies, including those for which we did not apply our weighted average approach. It should be noted that with empirical Bayesian estimation and with our approach, there are alterative means to generating prior probability distributions, such as subject matter expert (SME) judgment, that could be used to inform the prior distributions for the additional effects in Appendix C. This was considered to be beyond the scope of this study, and is suggested for future research.

When one source reported multiple effects for a particular predictor and team performance relationship in an analog environment, we calculated a linear composite to represent the effect for that particular source (Hunter & Schmidt, 2004). We corrected the observed correlations in a given analog study for unreliability of the predictor and criterion in order to match the corrections used in meta-analyses that were used to inform our prior distribution. Although we would have preferred not to correct the local validity estimates for unreliability because of the small sample sizes on which they were based, the majority of the variances used to inform the prior distributions were corrected for unreliability. Newman et al. (2007) indicate the importance of ensuring that the prior and local effects have the same corrections. When reliability was not reported, we used the closest estimation of reliability from the most similar research in our data set. When the correction resulted in an estimate greater than 1, we did not compute a weighted average. This is because our approach relied on the z transformation, which for values over 1 is undefined. Values exceeded 1 for correlations from Gunderson and Nelson (1963) and Gunderson and Ryman (1967), which were based on the same source data (e.g., self-report cooperation and performance) and exceeded .90 prior to correction. In the next section, we present the results of our weighted averages approach to estimating the relationship between team functioning and team performance in LDSEM-analog environments.

RESULTS

WEIGHTED AVERAGES

Figure 1 summarizes the weighted averages results of the team factor and team performance relationship in the LDSEM-analog environment, as well as the credible intervals around the estimates. Specific information about the local validity information obtained from LDSEM-analog studies, the meta-analytic effects that we used in the calculation of the weighted averages, and the estimated posterior distributions are provided in Table 1. Figure 1 displays the estimated results in a forest plot. Local validity estimates include team performance with cohesion, age homogeneity, education homogeneity, team learning, planning, team task-relevant experience, cooperation, and transformational leadership.

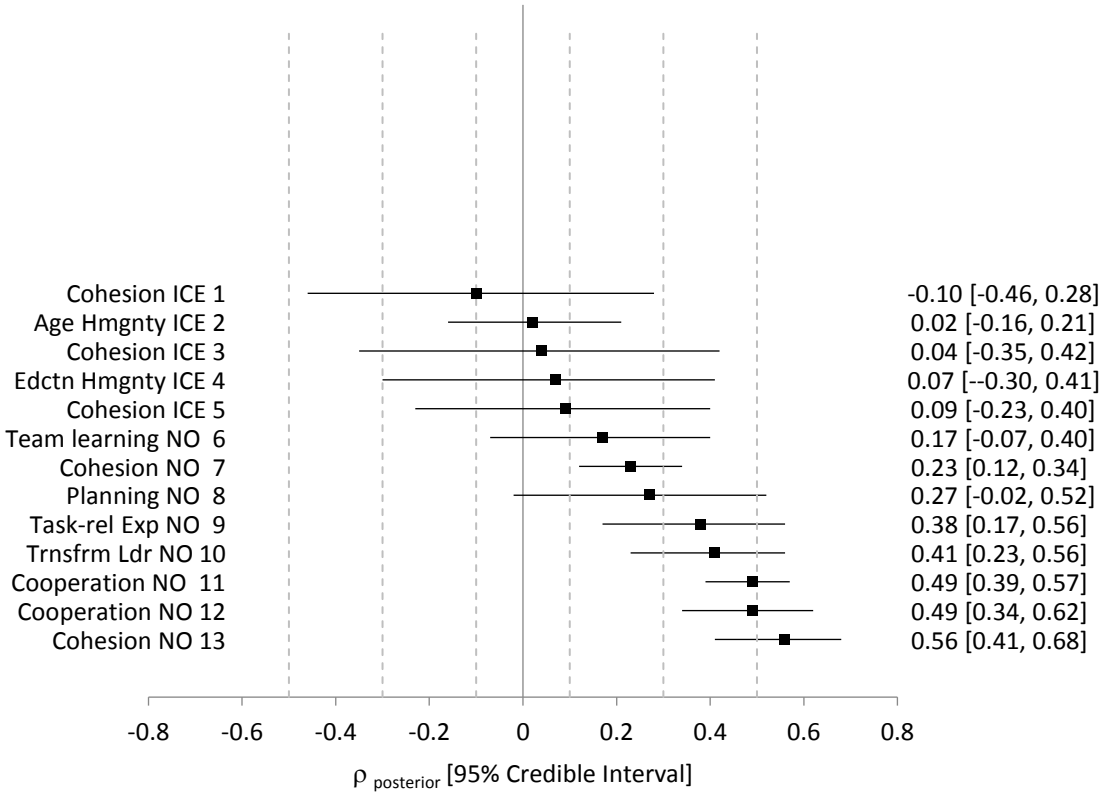


Figure 1. **Estimated distributions for the predictor and team performance relationships in analog environments.**

Figure note. ICE = analog team was in an ICE environment (e.g., winter-overs in Antarctica) NO = team was living and working together but not in an ICE. Hmgnty = Homogeneity. Task-rel Exp = task-relevant experience. Trnsfrm Ldr = Transformational Leadership. The square represents the weighted average local validity population estimate ($\rho_{posterior}$) and the bar represents the 95% credible interval. Specific estimates are provided in the right column as $\rho_{posterior}$ [95% credible interval]. The credible intervals can be interpreted as follows: there is a 95% chance that the true population predictor and team performance relationship (ρ) is between the first number and the second number. Numbers in the left column indicate the analog data source. 1 Gunderson & Nelson (1963), outcome = self-report team achievement, Antarctic stations; 2 & 4 from Gunderson & Ryman (1967), outcome = team accomplishment, mixed sources, Antarctic stations; 3. Emurian & Brady (1984), outcome = performance on lab task; 10-day isolation lab experiment; 5. Nelson (1964), outcome = supervisor ratings of individual performance aggregated within station, Antarctic stations; 6, 8, & 9 Knight (2015), outcome = team's time and number of obstacles completed in a final challenge task, military training; 7 & 11 Kniffin, Wansink, & Devine, (2015), outcome = supervisor rating of performance, firehouses; 10, 12, 13. Ko (2005) outcome = team performance, mixed sources, special operations teams.

TABLE 1 - ANALOG DATA SOURCES AND META-ANALYTIC EFFECTS CONTRIBUTING TO THE WEIGHTED AVERAGES

	Analog Data Source						Meta-Analysis Source and Relationship					
Fig. 1 #	Authors and Year		Performance correlate	*n*	*r* or rank	ρ_{local}	Meta-analysis used	Meta-analytic relationship used	ρ_{prior}	$\hat{\sigma}^2_{\rho_prior}$	*k*	$\bar{n}_{meta}$
1	Gunderson & Nelson	1963	Cohesion	9	-0.59	-0.70	Beal et al. (2003)	Interpersonal attraction and outcome performance	0.14	0.05	31	46.65
2	Gunderson & Ryman	1967	Age Hmgnty	15	-0.01	-0.02	Bell et al. (2011)	Age hmgnty and team performance	0.03	0.01	40	273.83
3	Emurian & Brady	1984	Cohesion	7	-0.41	-0.41	Beal et al. (2003)	Interpersonal attraction and outcome performance	0.14	0.05	31	46.65
4	Gunderson & Ryman	1967	Education Hmgnty	15	0.13	0.14	Bell et al. (2011)	Educational level hmgnty and general performance	0.01	0.06	14	279.57
5	Nelson	1964	Cohesion	14	-0.05	-0.06	Beal et al. (2003)	Overall cohesion and effectiveness	0.16	0.04	40	47.48
6	Knight	2015	Team learning	33	0.02	0.02	Lepine et al. (2008)	Overall team process and team performance	0.31	0.03	40	78.13
7	Kniffin et al.	2015	Cohesion	244	0.20	0.24	Beal et al. (2003)	Interpersonal attraction and outcome performance	0.14	0.05	31	46.65
8	Knight	2015	Planning	33	0.25	0.26	Lepine et al. (2008)	Transition and team performance	0.29	0.07	26	70.69
9	Knight	2015	Task-relevant experience	33	0.45	0.45	Bell (2004)	Team task-relevant expertise and team performance	0.33	0.02	9	57.44
10	Ko	2005	Trnsfrm Ldrshp	89	0.38	0.42	Wang et al. (2011)	Transformational leadership and team performance	0.33	0.07	34	83.24
11	Kniffin et al.	2015	Cooperation	244	0.43	0.51	Lepine et al. (2008)	Overall team process and team performance	0.31	0.03	40	78.13
12	Ko	2005	Cooperation	89	0.49	0.55	Lepine et al. (2008)	Overall team process and team performance	0.31	0.03	40	78.13
13	Ko	2005	Cohesion	89	0.56	0.63	Beal et al. (2003)	Interpersonal attraction and outcome performance	0.14	0.05	31	46.65

Note. Hmgnty = Homogeneity. Trnsfrm Ldrshp = Transformational Leadership. *n* = sample size of the analog study; *r* or rank = effect size taken from the analog study, or a linear composite or relevant effect sizes from the study; ρ_{local}= *r* from the analog study corrected for predictor and criterion unreliability; ρ_{prior} = meta-analytic mean corrected correlation (estimate chosen was corrected for predictor and criterion unreliability); $_{\rho_prior}$ = the estimate from the meta-analysis of the true population variability in effect sizes; *k* = the number of primary studies included in the meta-analysis; n_{meta} = average sample size of a primary study included in the meta-analysis.

First, we discuss the team cohesion and team performance relationships. Studies 1, 3, and 5 (as noted in Figure 1) were conducted in ICE; each of these studies measured team cohesion and team performance with different operationalizations. Estimate 1 reflects the estimated results for the team cohesion and team performance relationship for data collected in Antarctic stations where team cohesion was operationalized as self-rated compatibility of station members and team performance was operationalized as self-rated station achievement. The mean estimated validity is -.10, and with 95% certainty, we estimate that the true population validity falls between -.46 and .28. This is rather imprecise, as the prediction interval includes large, moderate, and small negative effects, no effect, and small and moderate positive effects. For estimates 3 and 5, the mean estimated validity is positive, but the credible interval suggests imprecision in the estimate; again, there is a wide range of possible effects within the prediction interval. Estimate 3 reflects the relationship between time spent socializing and performance on a lab task for data collected in a 10-day programmed environment (Emurian & Brady, 1984). Estimate 5 reflects the team cohesion and team performance relationship for station members in the Antarctic where cohesion was operationalized as the ratio of in-group members to total members in Antarctic stations and team performance was operationalized as individual-level performance appraisals aggregated to the station level.

Estimates 7 and 13 reflect the team cohesion and team performance relationships for teams that are sometimes used as LDSEM-analogs but which are not isolated or confined for extended periods. Estimate 7 represents the relationship between team cohesion and team performance in firehouses, where team cohesion was operationalized as spending time together (i.e., cooking together and eating together) and team performance was operationalized as supervisor ratings of global team performance. The mean estimated validity can be characterized as a small to medium effect (.23), and with 95% certainty the true population validity is estimated to fall between a small (.12) and medium-sized (.34) effect. Estimate 13 represents the team cohesion (self-report) and team performance (multiple raters) relationship in special operations military teams. Estimation results suggest, with 95% certainty, that the relationship is positive and moderate to large .56 (.41, .68).

In sum, we cannot, with 95% certainty, speak to the direction or size of the team cohesion and team performance effect in settings that may be most similar to LDSEM (e.g., ICE analogs). Conversely, with 95% certainty, we can describe the team cohesion and team performance relationships in the firehouses studied as positive and small to moderate, and in the special operations military teams studied, as positive and moderate to large.

Data for a few additional relationships other than team cohesion and team performance were also available. The age homogeneity and team performance (Figure 1, Estimate 2) and the educational level homogeneity and team performance relationships (Figure 1, Estimate 4) in an ICE setting (e.g., Antarctic station winter-over) were estimated with a large degree of imprecision; the prediction interval included positive, negative and no effect. Conversely, with 95% certainty, the true population effect between cooperation and team performance is positive and large (Figure 1, Estimates 11, 12) for firehouses and special operations teams. Finally, with 95% certainly, the true population effect between transformational leadership and team performance for special operations teams, and the true population effect between team task-relevant expertise and team performance for military training teams are positive and exceed a small effect (Figure 1, Estimates 9, 10).

Taken together, there is a high degree of imprecision associated with estimates of the true predictor and team performance relationships in ICE settings. Specifically, unlike most of the

estimated relationships in non-ICE settings, if we retain a 95% level of certainty, we have limited to no understanding of the size or direction of the relationship of team cohesion and team performance observed in multiple ICE environments, age homogeneity and team performance in ICE, and educational homogeneity and team performance in ICE, given the current available data from which effect sizes can be generated.

FIGURES BENCHMARKING TEAM FUNCTIONING OVER TIME

Next, we benchmarked team functioning in studies with sample sizes too small to generate a between-team effect size. We coded studies with fewer than 5 teams using a similar coding sheet in terms of moderators, and coded a team-level representation of variables of interest (e.g., team mean) such as performance over time. We plotted team functioning over time when data were comparable (e.g., similar scales, similar response formats), reported for at least three different teams, and from at least two different data sources (e.g., articles, conference presentations). We reviewed the research and quantified data where possible to help inform our recommendations for future research. Figures on performance are presented in terms of efficiency, team conflict, communication with MC, and team mood. Additional constructs measured in the benchmarking data for which we did not have comparable published data are listed in Appendix C, Table 3[2].

Performance Over Time

Homeostat was used to collect data on team performance across a number of space simulations (e.g., Human Behaviour in Extended Spaceflights [HUBES], SFINCSS). Homeostat is a computer task in which, under time pressure, a team solves tasks that require the coordinated action of the whole team. A number of metrics can be assessed using Homeostat, including an efficiency metric (Csh) and leadership tactics. Two key studies (e.g., Eskov, 2011; Vinokhodova et al., 2001) reported Csh results across multiple teams, which allowed us to plot team efficiency over time. Figure 2a is a plot of team efficiency across mission days. Figure 2b is a plot of team efficiency over relative time (i.e., the mission day divided by the total mission length).

The data suggest that three teams (i.e., a team in the EXEMSI and two of the teams in SFINCSS simulations) were relatively consistent in terms of efficiency over time. The HUBES team decreased steadily in efficiency over time. One of the SFINCSS teams (Group 3) had a sharp decline in efficiency early in the simulation and then steadily increased during the remainder of the simulation.

Descriptive information on team functioning in the HUBES and SFINCSS simulations implicate ineffective role structure and conflict as possible triggers of the performance decrements of HUBES and SFINCSS - Group 3. Specifically, in addition to measures of efficiency, the Homeostat also collects information on leadership tactics by individual team members as a means of understanding the leadership structure used while completing the task. Vinokhodova, Bystritskaya, and Eskov (2002) provided a detailed analysis of SFINCSS Homeostat data and describe the three SFINCSS groups as follows (p. 107).

---

[2] We appreciate Dr. Pete Roma who shared unpublished team cohesion data collected across multiple simulations (e.g., HERA, NEEMO), and Dr. Kim Binsted who shared unpublished team mood data from HI-SEAS with us. Given that the data are unpublished and the level at which we present these figures (i.e., raw data across time), we do not report the figures in this report.

"The 240 day group had a fixed role distribution with one dominating leader. The 110 day Group 2 was characterized as a flexible role distribution with the possibility of exchanged leadership function between the three crew members. In contrast, for SFINCSS Group 3, the evident leader was absent: There were two subjects demonstrating leadership attempts. The capacity decreased for subject K at the end of isolation, but the increase in leadership activity for subject I wasn't sufficient in this period, so we cannot make a conclusion that the role distribution structure was sufficiently developed in this group."

The SFINCSS simulation also included a New Year's Eve incident between a member of another group and a woman in Group 3 of the simulation, which led to tension between crews (Sandal, 2004). The sharp decrease in effectiveness in the SFINCSS Group 3 depicted in Figure 2a also happened around this time. For HUBES, Sandal (2001) reports that there was evidence of an unstable crew structure; specifically, the commander's leadership was challenged during the first 8 to 10 weeks of the mission. Further, crew relations in the simulation were marked by interpersonal tension and alienation of one crew member during later parts of the experiment. Taken together, this may suggest ineffective role structure, conflict, and alienation as possible threats to team efficiency, all of which are worth additional research. We next explore team conflict, communication, and mood over time.

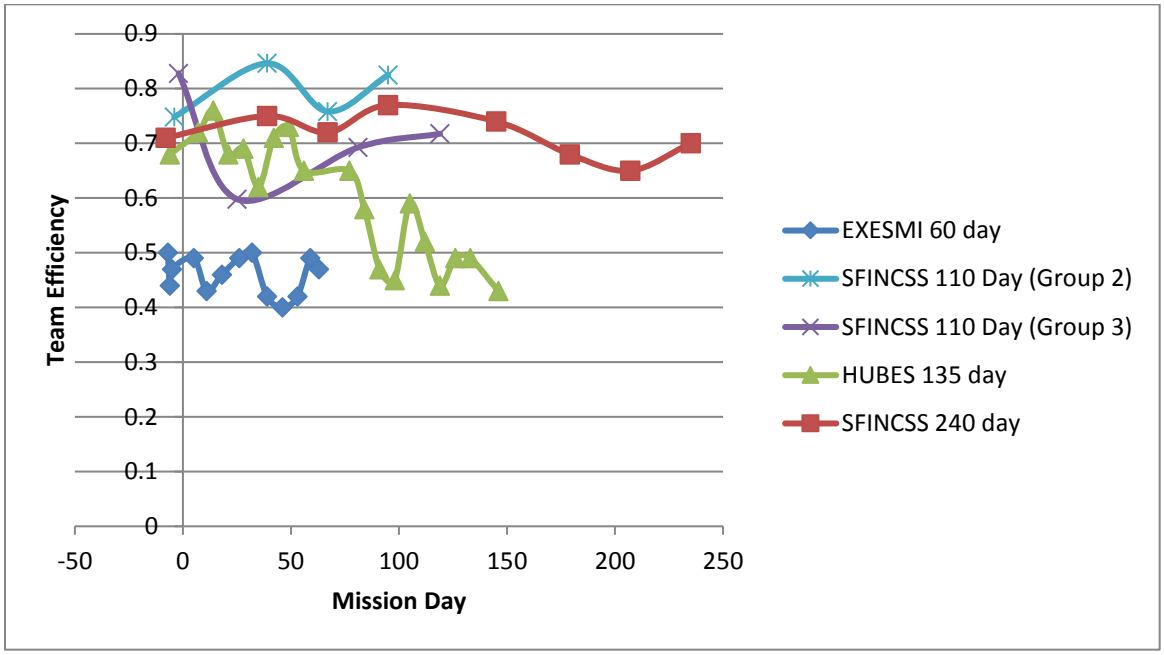


Figure 2a. **Team efficiency over time.**

Note. Figure includes Csh data collected using a Homeostat device reported in Eskov (2011; EXEMSI, HUBES, SFINCSS 240) and Vinokhodova, Bykstritskaya, & Eskov (2001) (SFINCSS 110-day missions).

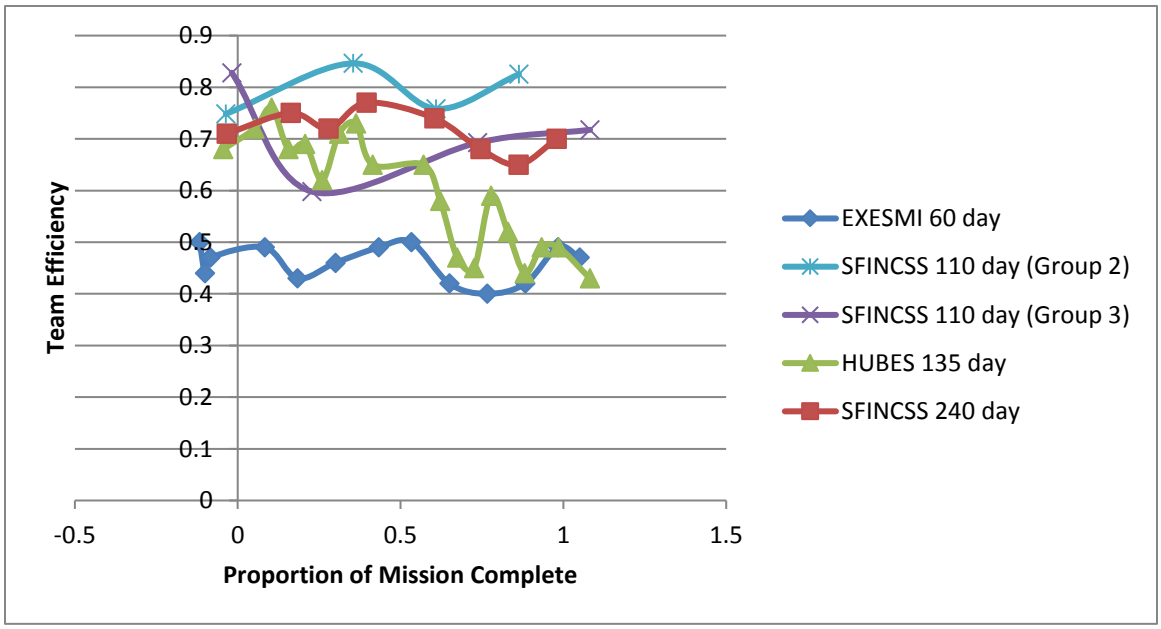


Figure 2b. **Team efficiency over relative time.**

Note. Figure includes Csh data collected using a Homeostat device reported in Eskov (2011; EXEMSI, HUBES, SFINCSS 240) and Vinokhodova, Bykstritskaya, & Eskov (2001; SFINCSS 110). Relative time was calculated as the mission day divided by the total mission length.

Team Conflict Over Time

A few sources (k = 5) reported conflict scores over time for 8 different teams using 2 types of conflict metrics (e.g., total number of conflicts reported, Likert scale). Data were comparable across multiple teams from different analog environments for the total number of conflicts reported. Figures 3a and 3b show the number of conflict incidents at base during three different winter-overs at Scott Base in Antarctica (all reported in Natani, Shurley, & Joern, 1973) and Mars 500 (reported in Basner et al., 2015). Basner et al. (2015) reported that the relatively low level of within-crew conflict for the Mars 500 crew might, in part, be due to the conflict between the crew and MC. Figures 4a and 4b show the same data as Figures 3a and 3b, but also include the instances of conflict between the Mars 500 crew and MC as reported in Basner et al. (2015).

Data do not show a consistent trend across teams. Some teams are more variable than others in the number of conflict incidents per month, while others are more stable. Some teams report conflict early on, while others do not. By 40% of the mission completion (with this data the equivalent of at least 90 days) all teams had reported a least one instance of conflict either within base or between the crew and MC. No team had more than 6 instances of conflict per month with a given target (either within the crew or with MC).

[Remainder of this page intentionally left blank.]

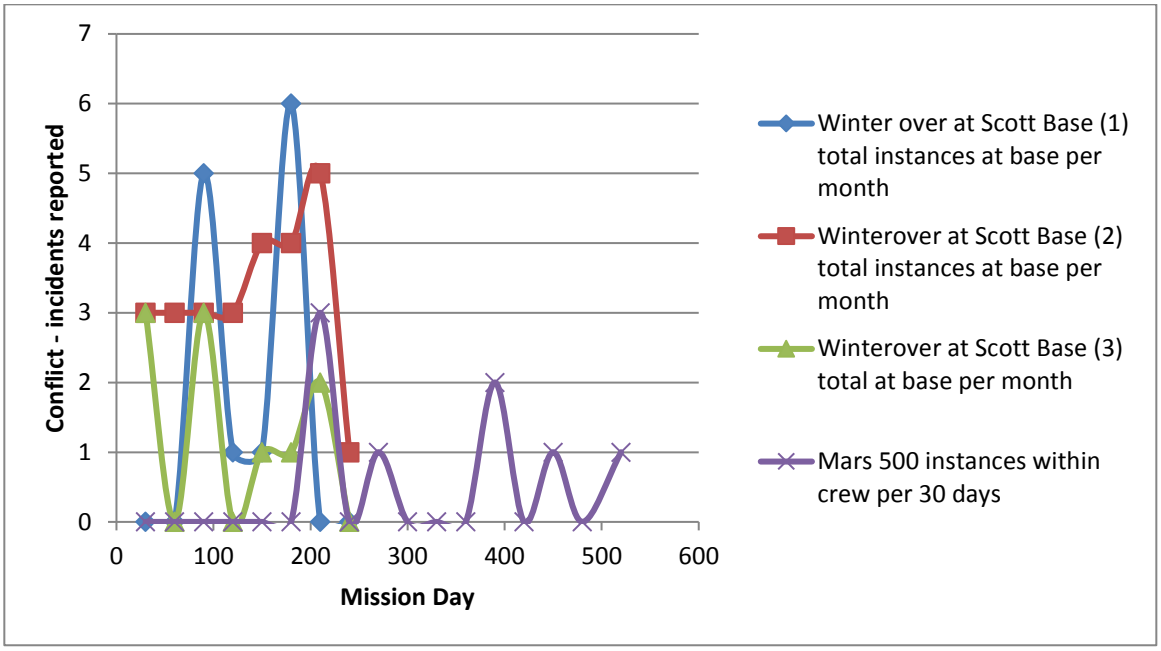


Figure 3a. **Number of perceived within-crew conflict incidents reported over time.**
Note. Figure includes data from Natani, Shurley, & Joern, (1973; Scott Base) and Basner, Dinges, & Mollicone (2014; Mars 500). Data instances are for every 30 days; for example, the 4th 30-day period for the first winter-over team at Scott Base reported 5 conflicts during that 30-day period.

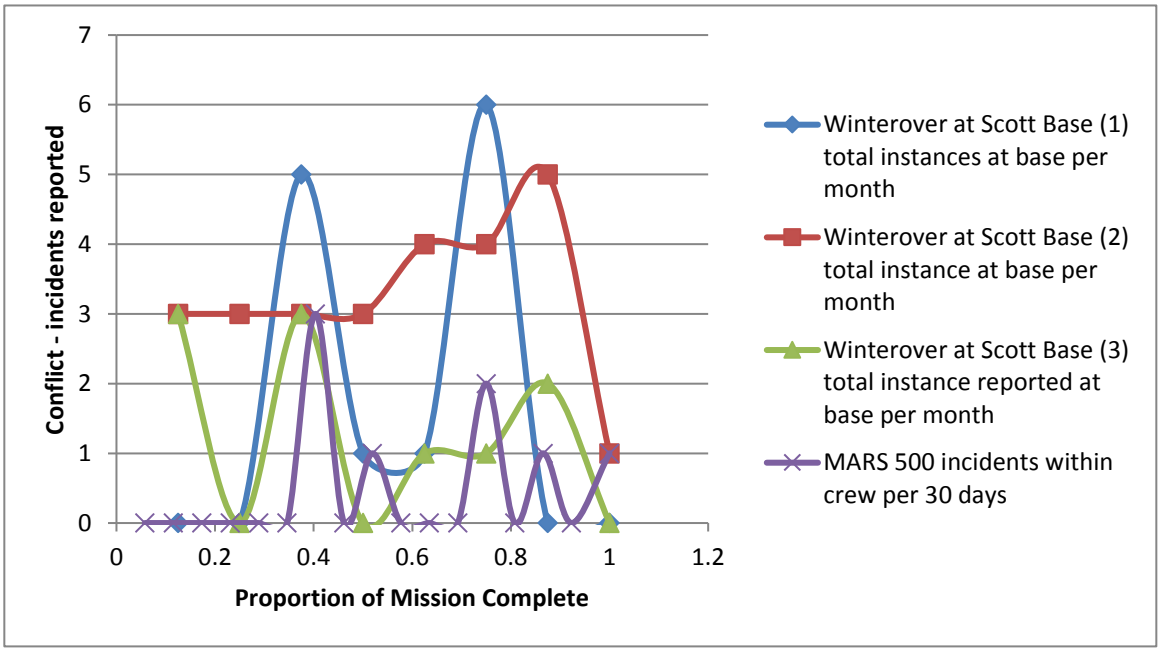


Figure 3b. **Number of within-crew conflict incidents reported over relative time.**
Note. Figure includes data from Natani, Shurley, & Joern, (1973; Scott Base) and Basner, Dinges, & Mollicone (2014; Mars 500). Data instances are for every 30 days; for example, the 4th 30-day period for the first winter-over team at Scott Base reported 5 conflicts during that 30-day period. Relative time was calculated as the mission day divided by the total mission length.

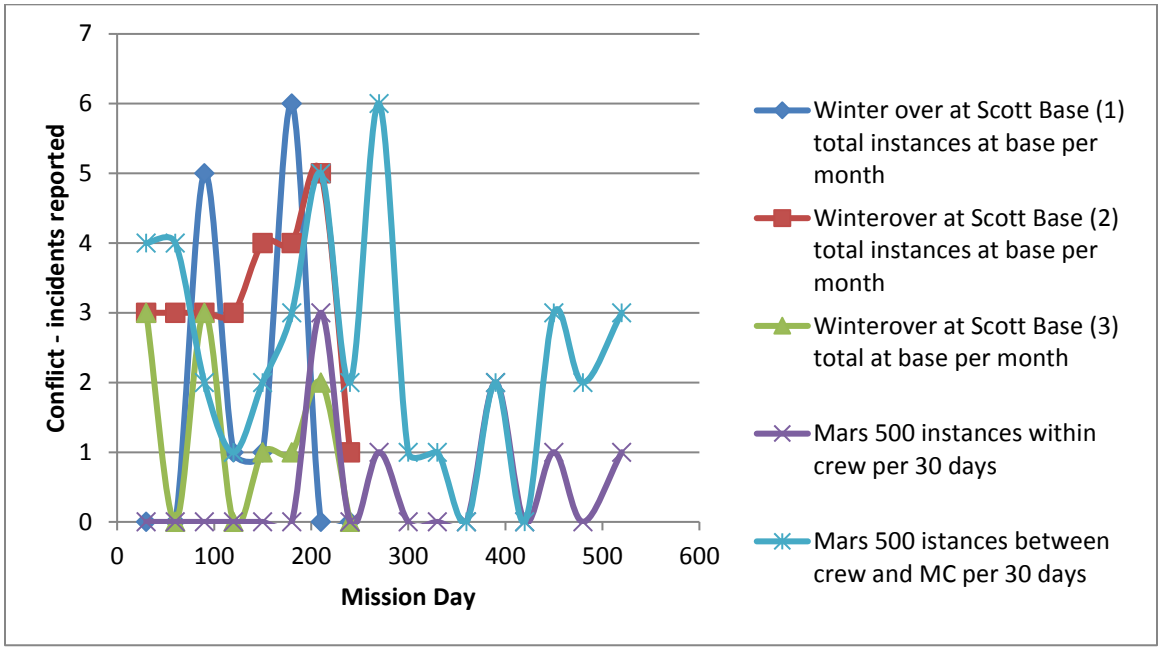


Figure 4a. **Number of conflict incidents reported over time.**

Note. Figure includes data from Natani et al. (1973; Scott Base) and Basner et al. (2014; Mars 500). Figure is similar to Figure 4a, but also includes Mars 500 conflict instances with MC. Data instances are for every 30 days; for example, the 4th 30-day period for the first winter-over team at Scott Base reported 5 conflicts in that 30-day period.

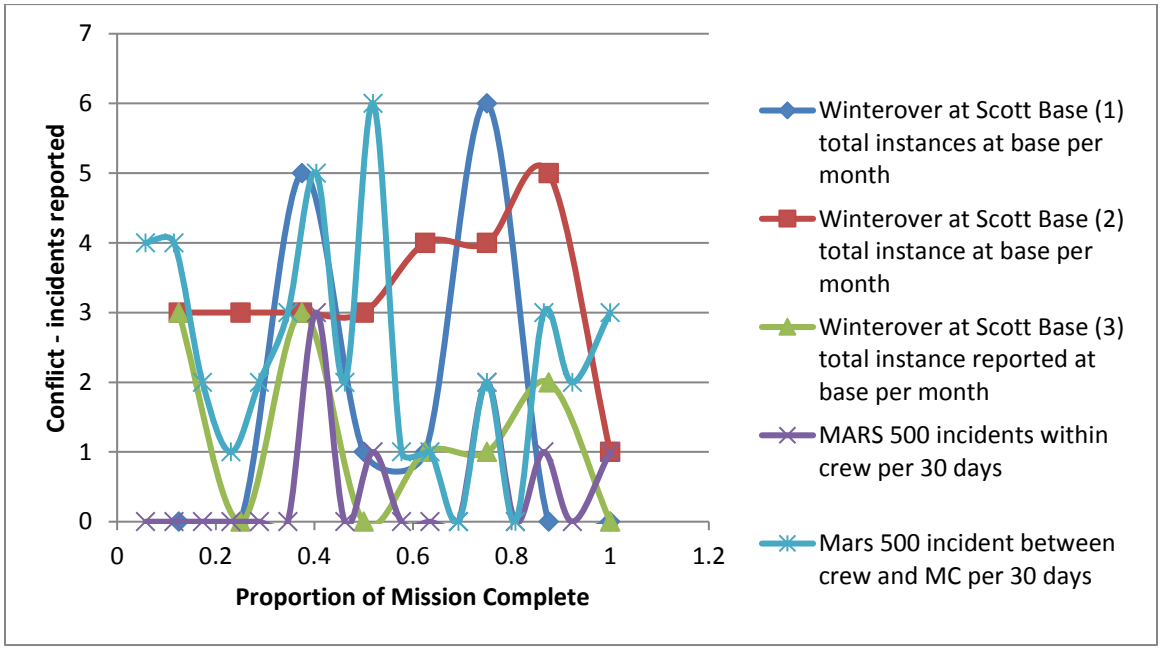


Figure 4b. **Number of conflict incidents reported over relative time.**

Note. Figure includes data from Natani et al. (1973; Scott Base) and Basner et al. (2014; Mars 500). Figure is similar to Figure 4b, but also includes Mars 500 conflict instances with MC. Data instances are for every 30 days; for example, the 4th 30-day period for the first winter-over team at Scott Base reported 5 conflicts in that 30-day period. Relative time was calculated as the mission day divided by the total mission length.

COMMUNICATION WITH MISSION CONTROL

Next we explored communication. Communication between space crews and MC is thought to provide information about the crew's psychological health and the crew's psychological climate. Analysis of a space crew's communication with MC is the standard operating procedure of the psychological support group in Russian MCs (Gushin et al., 2012). Specifically, the psychological support group content analyzes crew-MC interactions to examine crews' emotional status and the communicators' coping strategies (Gushin et al., 2012).

Gushin and colleagues have examined crew communication with MC in several studies (e.g., Gushin et al., 1997; Gushin et al., 2001; Gushin & Yusupova, 2003) and have reported comparable data, which allowed us to plot the total duration of crew-MC audio-communication sessions (in seconds) over time (see Figures 5a and 5b), as well the average report length per week of the commander's end-of-day report to MC (see Figures 6a and 6b). For the SFINCSS, HUBES, and ECOPSY simulations, audio communication paralleled the standards of Mir in that 30 minutes were made available for audio communication every 90 minutes in the daily schedule but use of the time was not required. At the end of each day, the commander submitted a written report to MC on mission status and fulfillment of the daily schedule (Gushin et al., 1997; Gushin et al. 2001). Data in Gushin and Yusupova (2003) was collected by researchers listening to crew-MC communication once a week (for ISS mission 1) and twice a week (for ISS mission 2).

As depicted in Figures 5a and 5b, patterns of average audio-communication length between the commander and MC were inconsistent across teams. It is interesting to note, that the HUBES crew that had decreasing efficiency over time (Figure 2a) also had shorter audio communication with MC over time (Figures 5a and 5b). As depicted in Figures 6a and 6b, average mission report length to MC per week decreased over the course of the mission in SFINCSS, EXEMSI, and ECOPSY. Gushin et al. (2012) describe this as the closing of a communication channel, or psychological closing. Psychological closing can include a decrease of the communication volume throughout isolation, decrease in the issues discussed, and preference for communication partners.

It should be noted that there is a wealth of specific details (e.g., negative statements, jokes) that can be gleaned and assessed via content analysis of within- and between-group communications. Our figures here only reflect report length and total time for audio communication, which were reported in the same format across multiple teams. We refer the interested reader to Gushin et al. (2012) and Tafforin, Vinokhodova, Chekalina, and Gushin (2015) for more detail on the range of communication parameters that have been examined.

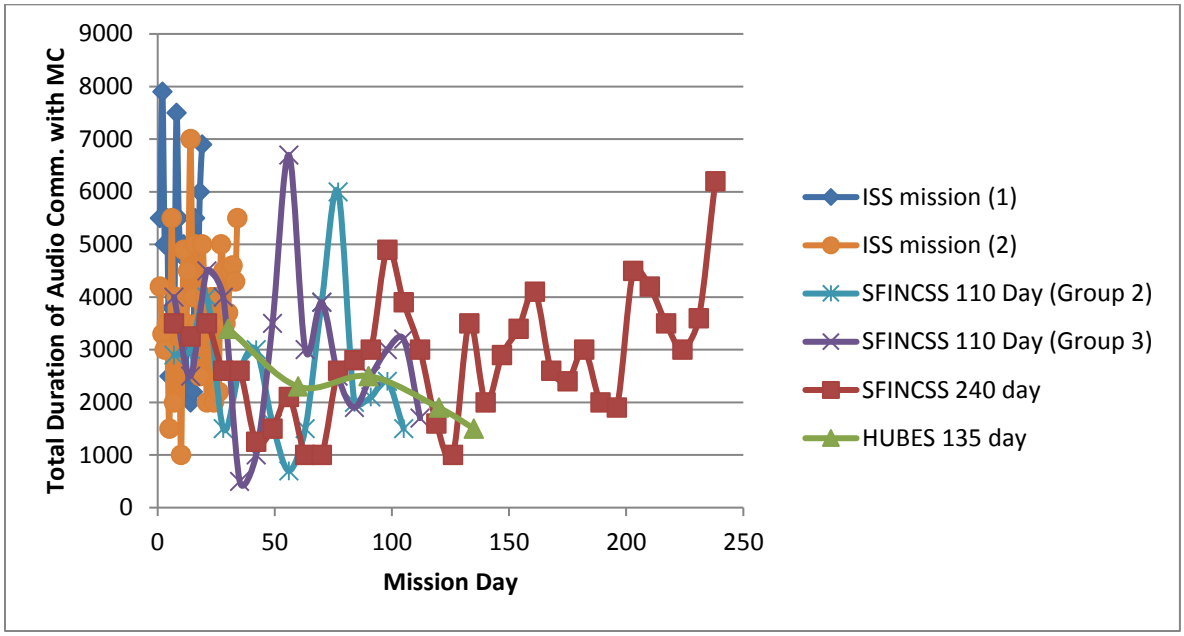


Figure 5a. Total duration of audio-communication MC over time

Note. Comm. = communication; MC = mission control. Figure includes data from Gushin & Yusupova (2003; ISS); Gushin et al. (2001; SFINCSS 110 and 240); Gushin et al. (1997; HUBES)

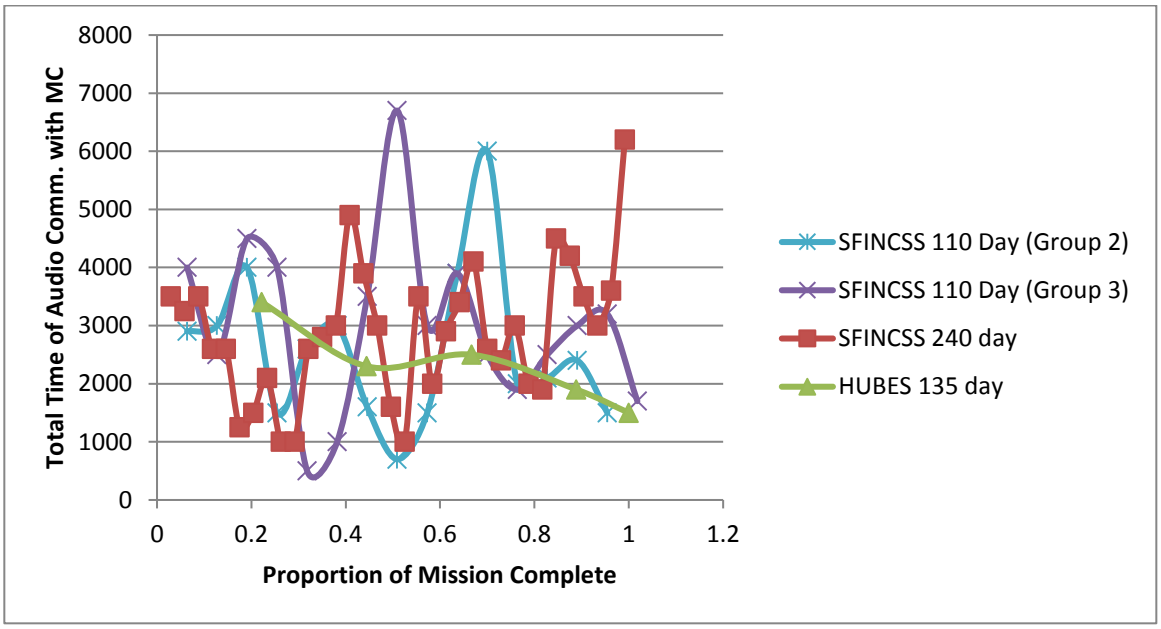


Figure 5b. Total duration of audio-communication with MC over relative time

Note. Comm. = communication; MC = mission control. Figure includes data from Gushin et al. (2001; SFINCSS 110 and 240); Gushin et al. (1997; HUBES). Relative time was calculated as the mission day divided by the total mission length.

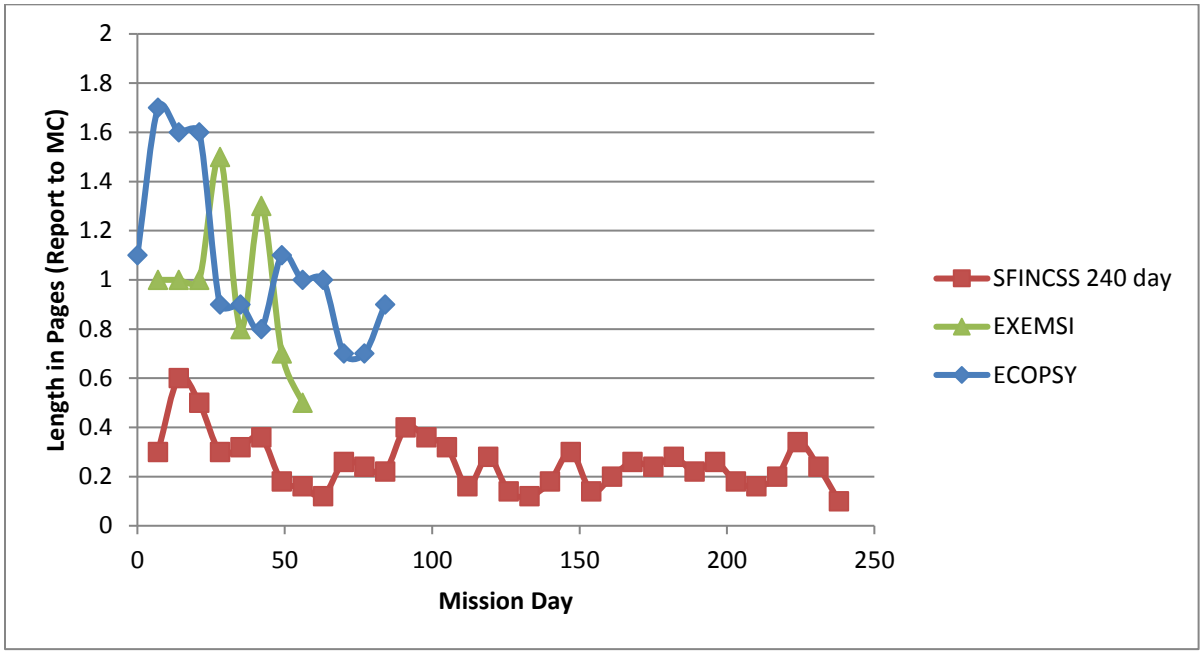


Figure 6a. Average mission report length to MC per week over time

Note: Figure includes data from Gushin et al. (2001); Gushin et al. (1996; EXEMSI); Gushin, Zaprisa, et al. (1997; ECOPSY).

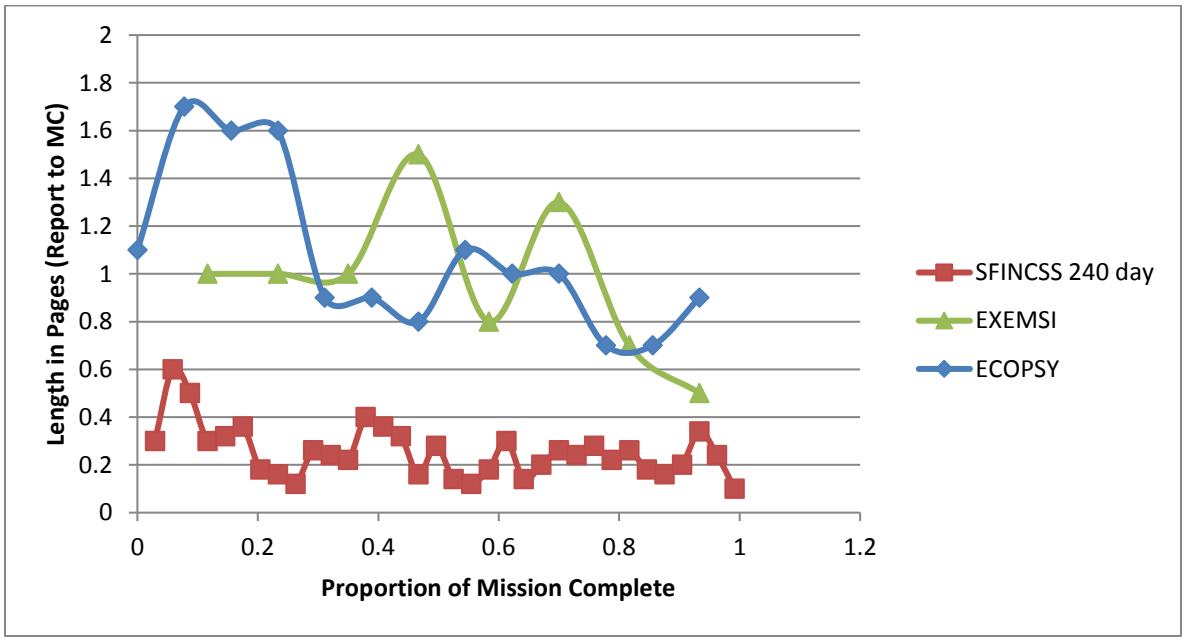


Figure 6b. Average mission report length to MC per week over relative time

Note: Figure includes data from Gushin et al. (2001); Gushin et al. (1996; EXEMSI); Gushin, Zaprisa, et al (1997; ECOPSY). Relative time was calculated as the mission day divided by the total mission length.

Team Mood

Though mood is typically measured in LDSEM-analog research as an individual-level variable, researchers sometimes use the team mean of individual-level mood scores to represent team mood. Team mood is important because it contributes to team emotion, which is defined as a team's affective state that arises from bottom-up components such as affective composition, and top-down components such as affective context (Kelly & Barsade, 2001). Team emotion starts with individual-level moods and emotions and is then shared with the team either implicitly through emotional contagion or explicitly through means such as affect management. Environmental context, such as lighting, physical layout, noise, and heat, can all affect moods (see Kelly & Barsade, 2001 for a review).

It is not surprising that mood has often been examined within the extreme context of LDSEM-analogs. Multiple studies reported the affect of team members using Profile of Mood States (POMS; Curran, Andrykowski, & Studts, 1995; Shacham, 1983). POMS captures individuals' mood via self-report ratings on six dimensions using a 5-point Likert scale. The dimensions are tension-anxiety, depression-dejection, anger-hostility, fatigue-inertia, confusion-bewilderment, and vigor-activity. To arrive at an overall total mood disturbance score, the first five subscales listed are summed and then the vigor-activity subscale is subtracted. Team mood is captured with the average total mood disturbance across the team. Figures 7a and 7b show team mood over time and team mood over relative time, respectively. Figure 7a shows that the MARS 500 crew reported elevated total mood disturbance compared with teams in other LDSEM-analog environments, although it should be noted that the scaling reported for Scott Base was 0 to 4 instead of 1 to 5 as in the other simulations. Thus, the winter-over at Scott Base may have ratings more similar to Mars 500. Both studies that included teams in ICE for a year or more (e.g., Mars 500, an Antarctic winter-over) showed a spike in team total mood disturbance around the 1-year mark, and this was confirmed in the text of the studies reporting the data (e.g., Steel, 2001; Wang et al., 2014). Figure 7b, which shows total mood disturbance over time relative to the proportion of the mission complete, does not support a clear third-quarter phenomenon at the team level.

Team mood also has been operationalized in LDSEM-analog environments as the team mean of self-report ratings on the positive and negative mood components of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988, see Leon and colleagues, 2004, 2011, for examples). Figures 8a, 8b, 9a, and 9b, show the relationship between affect operationalized as the team mean PANAS scores over time. Figures 8a and 9a show team positive affect over time and relative time. Figures 8b and 9b show team negative affect over time and relative time. For team negative affect over relative time, three of seven LDSEM-analog teams show an increased negative affect during the third quarter.

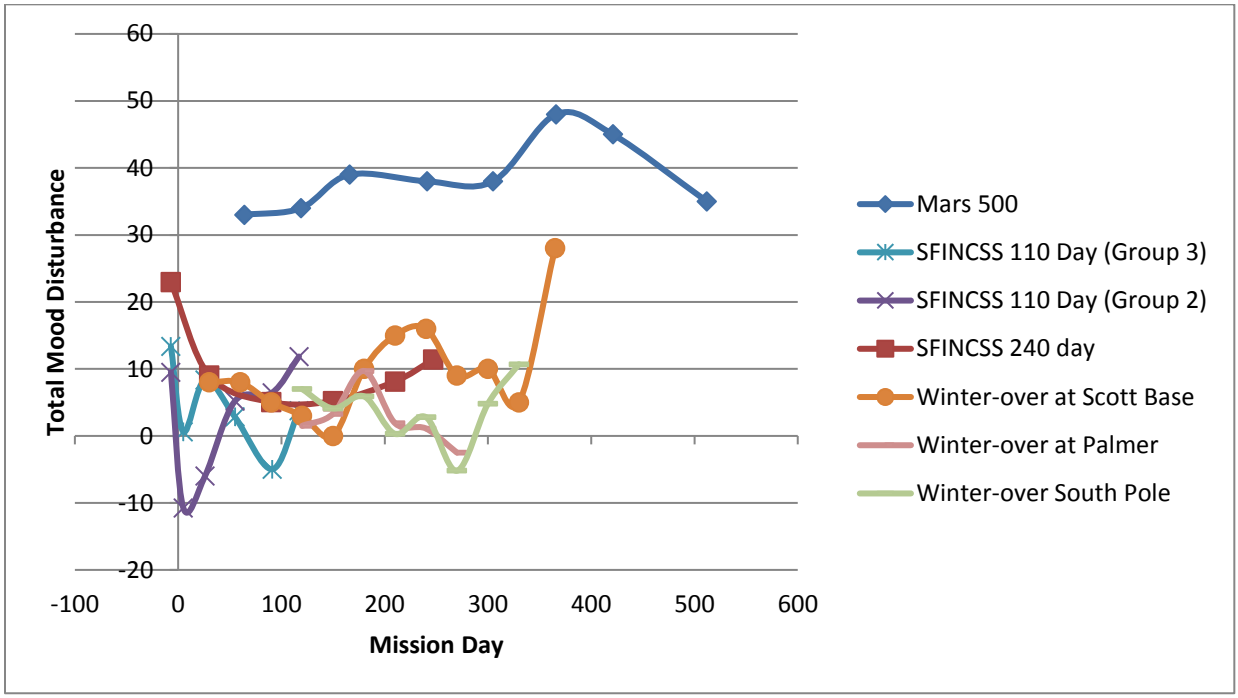


Figure 7a. Total mood disturbance (POMS) crew mean over time

Note. Figure includes data from Wang et al. (2014; Mars 500); Mizuno et al. (2001; SFINCSS 110 and 240); Steel (2001; Antarctic winter-over); Palinkas & Houseal (2000, Palmer and South Pole). Average total mood disturbance of the crew is plotted over time.

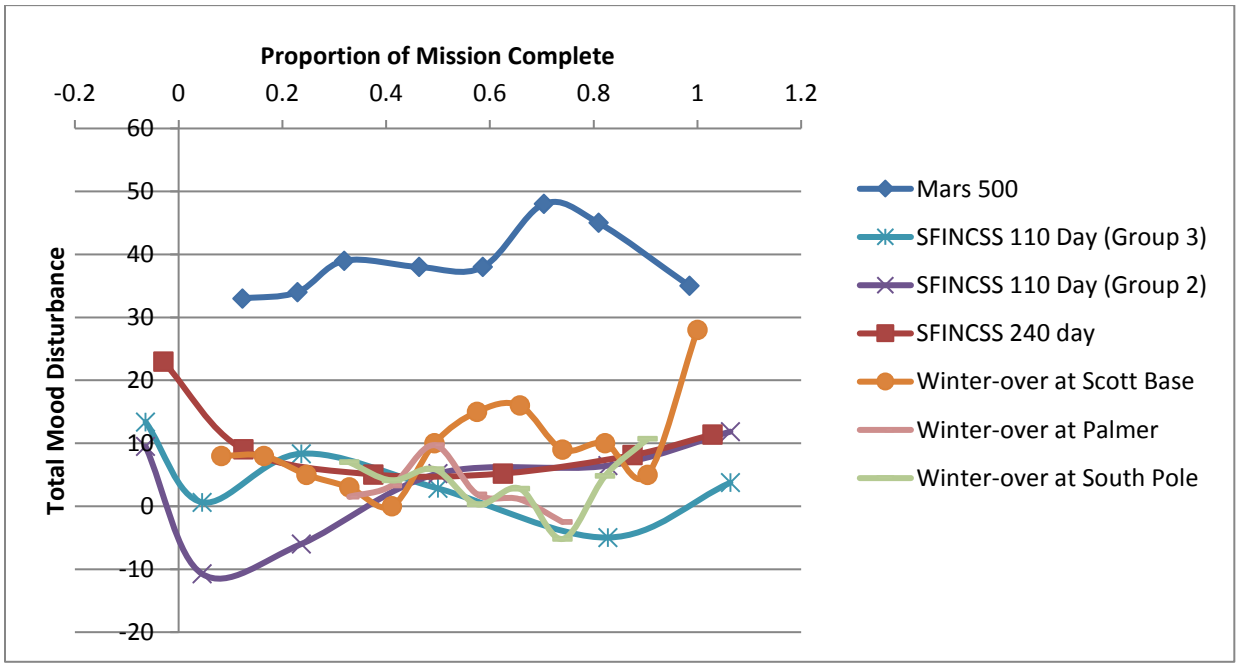


Figure 7b. Total mood disturbance (POMS) crew mean over relative time

Note. Figure includes data from Wang et al. (2014; Mars 500); Mizuno et al. (2001; SFINCSS 110 and 240); Steel (2001; Antarctic winter-over); Palinkas & Houseal (2000, Palmer and South Pole). Average total mood disturbance of the crew is plotted over relative time. Relative time was calculated as the mission day divided by the total mission length.

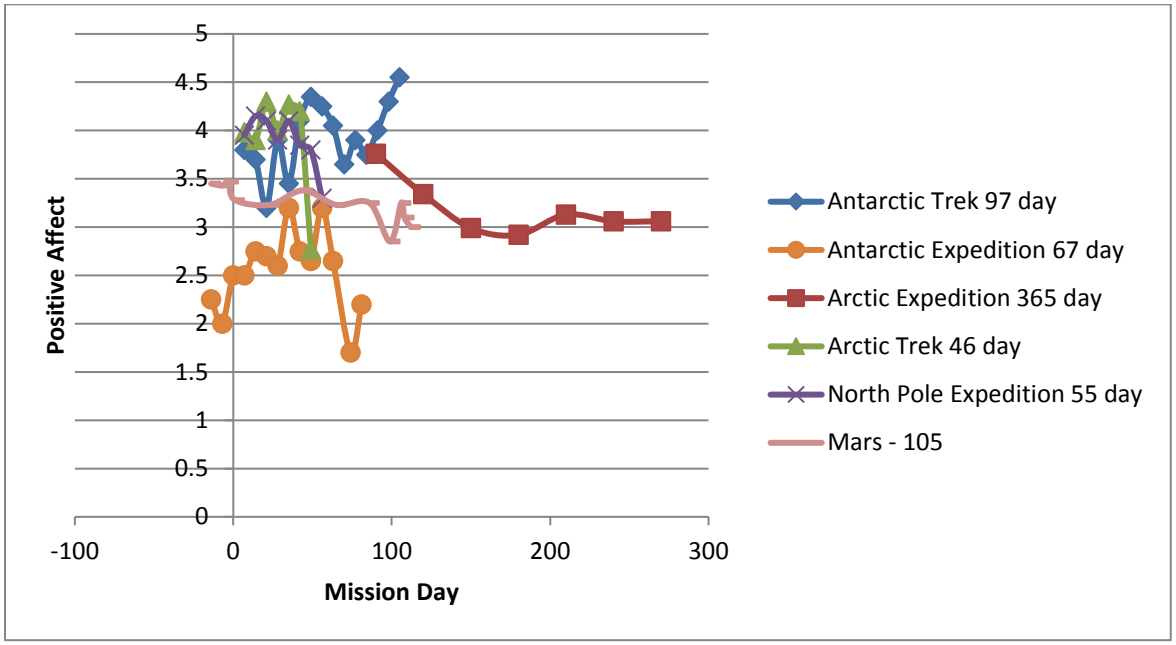


Figure 8a. Positive affect crew mean over time

Note. Figure includes data from Atlis et al. (2004; Antarctic Trek); Kahn & Leon (2000; Antarctic Expedition); Leon et al. (2002; Expedition 365); Leon et al. (2004; Expedition 46); Leon et al. (2011; Expedition 55); Nicholas et al. (2013; Mars 105); Data from Binsted (2015; HI-SEAS) were available but excluded from this version because they are unpublished.

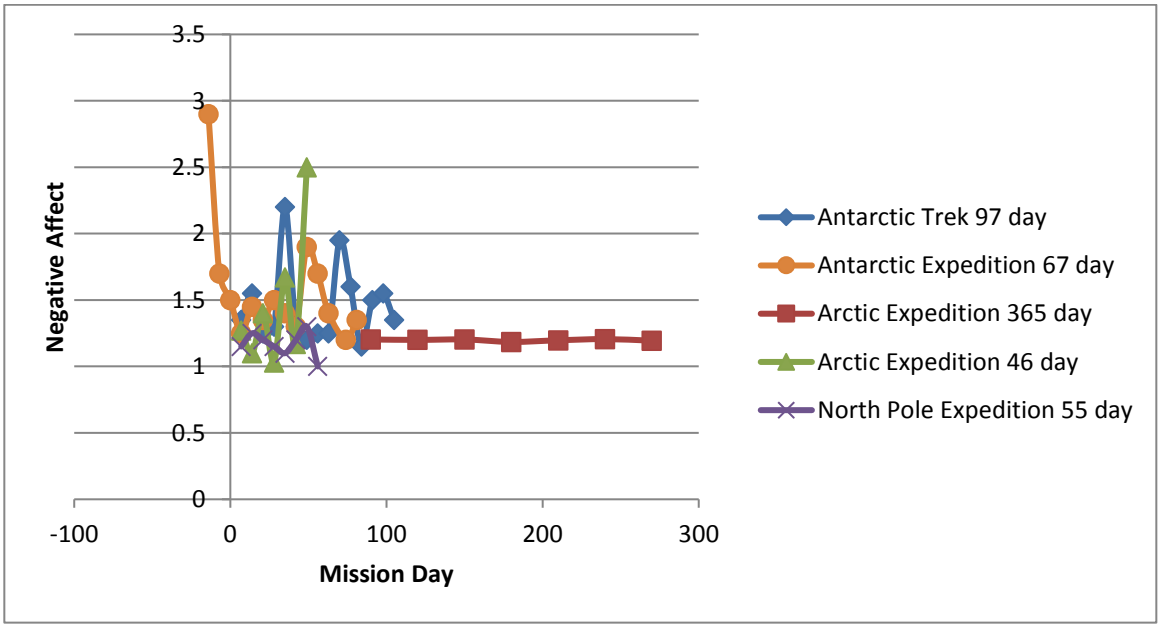


Figure 8a. Negative affect crew mean over time

Note. Figure includes data from Atlis et al. (2004; Antarctic Trek); Kahn & Leon (2000; Antarctic Expedition); Leon et al. (2002; Expedition 365); Leon et al. (2004; Expedition 46); Leon et al. (2011; Expedition 55); Nicholas et al. (2013; Mars 105); Data from Binsted (2015; HI-SEAS) were available but excluded from this version because they are unpublished.

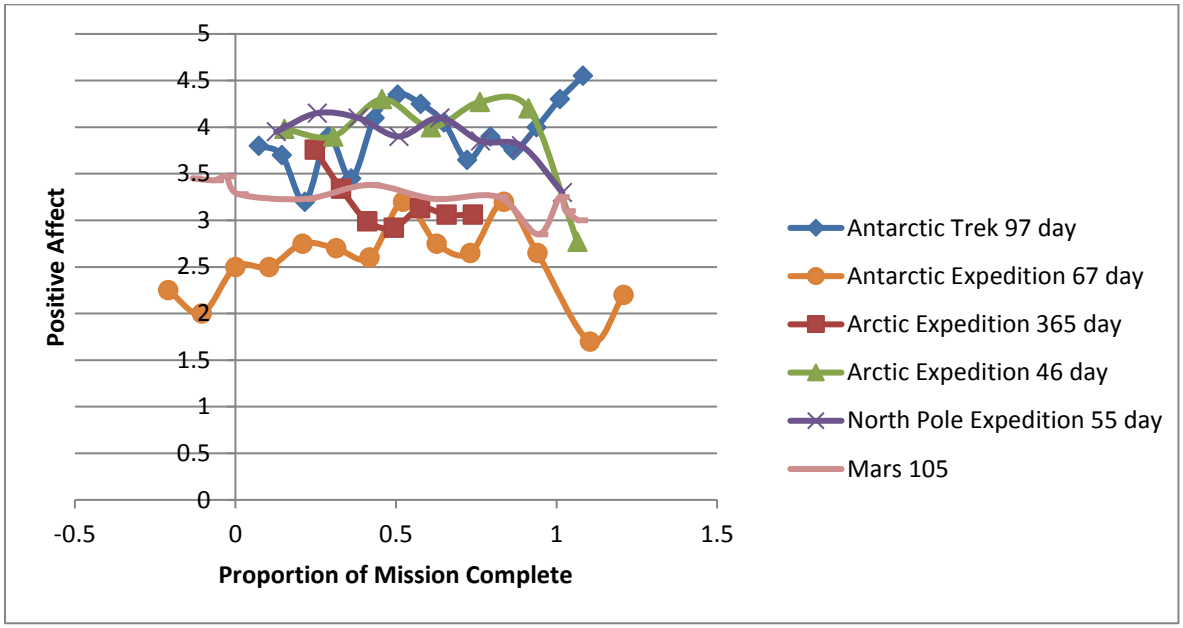


Figure 9a. Positive affect crew mean over relative time

Note. Figure includes data from Atlis et al. (2004; Antarctic Trek); Kahn & Leon (2000; Antarctic Expedition); Leon et al. (2002; Expedition 365); Leon et al. (2004; Expedition 46); Leon et al. (2011; Expedition 55); Nicholas et al. (2013; Mars 105); Data from Binsted (2015; HI-SEAS) were available but excluded from this version because they are unpublished. Relative time was calculated as the mission day divided by the total mission length.

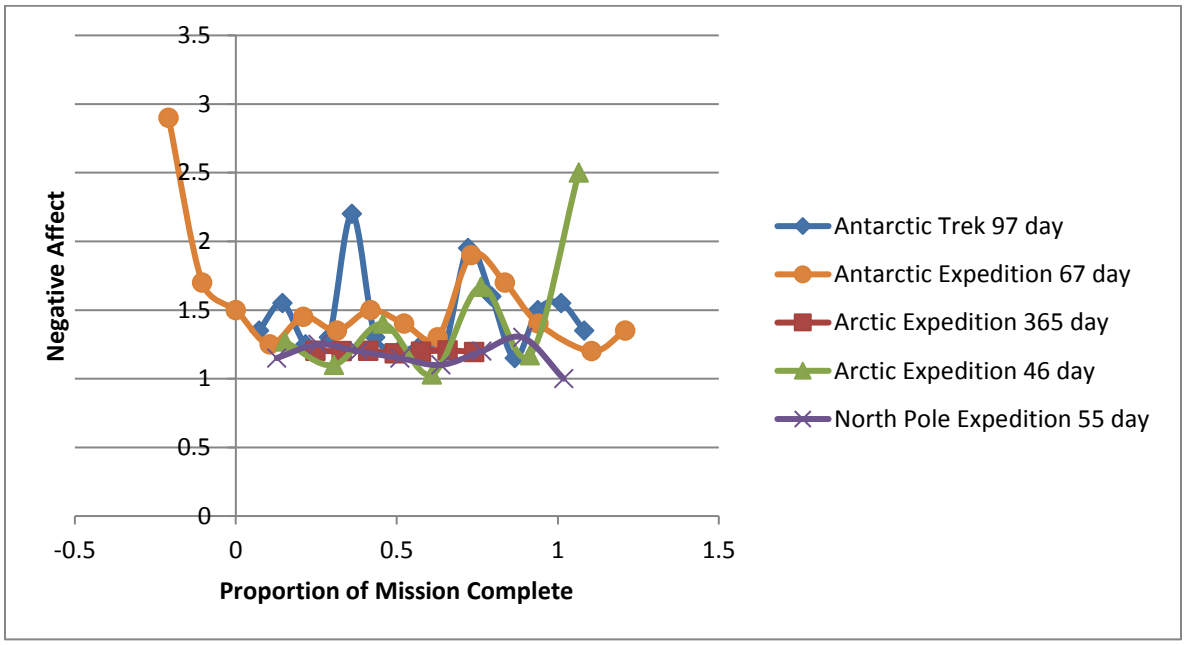


Figure 9b. Negative affect crew mean over relative time

Note. Figure includes data from Atlis et al. (2004; Antarctic Trek); Kahn & Leon (2000; Antarctic Expedition); Leon et al. (2002; Expedition 365); Leon et al. (2004; Expedition 46); Leon et al. (2011; Expedition 55); Nicholas et al. (2013; Mars 105); Data from Binsted (2015; HI-SEAS) were available but excluded from this version because they are unpublished. Relative time was calculated as the mission day divided by the total mission length.

GAPS AND RESEARCH RECOMMENDATION

In this section, we provide a list of gaps and research recommendations, along with a brief rationale for our recommendations. We have grouped our recommendations into three general areas. Recommendations regarding:

(1) collecting data in ICE environments such as the Antarctic and chamber simulations such as HERA,
(2) conducting research so that it better lends itself to quantitative aggregation, and
(3) gaps and research recommendations in regards to team factors in need of study.

Each of these sections includes a summary table that concisely lists specific research recommendations, in priority order where applicable.

COLLECTING DATA IN ICE ENVIRONMENTS SUCH AS THE ANTARCTIC AND CHAMBER SIMULATIONS SUCH AS HERA

1. The need for a better understanding of team functioning in isolated and confined settings.

The nature of the existing data (e.g., correlational, descriptive) does not allow us to make definitive statements about which specific characteristics (e.g., dangerous environment, isolation and confinement) impinge on team functioning. Intuitively, it seems that isolation and confinement is one of the most salient characteristics expected for LDSEM that could impinge on team functioning. The number of studies conducted in analog environments from which an effect size could be calculated was limited. The team cohesion and team performance relationship was the most commonly examined relationship in these studies. The available evidence on the team cohesion and team performance relationship suggests that we have little understanding of this relationship in teams that operate in isolated and confined settings. In contrast, the estimated results for the team cohesion and team performance relationship for teams that live and work together, but were not specifically in isolation and confinement (e.g., special operations teams, firefighters), had markedly smaller prediction intervals.

Some of the inaccuracy of the estimated results for the team cohesion and team performance relationships in ICE analogs was because the ICE environments had smaller sample sizes than the other analog teams: Smaller samples are associated with more sampling error for the local validity study. Even so, it should be noted that the raw local validity estimates between team cohesion and team performance in ICE setting were all negative (which was different than the positive effects observed in non-ICE analog settings). The discrepancy of the local effect from the meta-analytic effect also contributed to the imprecision of the estimated results of the weighted averages approach.

The weighted averages we report suggest that given the available data, we have no predictive accuracy for the team cohesion and team performance relationship in isolated and confined settings. The credible intervals included effects ranging from small to large and included both positive and negative effects. Conversely, our weighted averages for team performance with team cohesion, team cooperation, and transformational leadership in non-ICE analog teams show a different pattern. The estimation results indicated, with 95% certainty, the effects are positive, and in most cases the prediction intervals were relatively narrow. We recommend research on team functioning (e.g., team cohesion) specifically in analogs that have

teams in isolation and confinement (e.g., Antarctic stations, HERA environment) to better estimate the predictor and criterion relationships in this unique context relevant to LDSEM.

2. **The need for experimental and quasi-experimental studies to identify key threats to team functioning**.

The studies included in our review were almost exclusively correlational or descriptive (for an exception see the work by Emurian and colleagues). That means that we cannot make cause-and-effect statements for the relationships examined here, nor can we disentangle the effects of one predictor from another. We recommend that experimental and quasi-experimental designs (similar to the research conducted in HERA) be used where possible to identify key threats (e.g., task switching) to team functioning. Tying team functioning to specific threats can help NASA anticipate the need for and plan the efficient application of countermeasures during a team's life cycle.

3. **The need for research on crews that are in isolation and confined for long-durations.**

We provide four pieces of evidence in favor of long-duration analog studies. First, after examining Homeostat-Csh data across 4 space simulations, Eskov (2011) noted that the efficiency of group interactions did not change during the one-week CAPSUL experiment. Changes were observed in the other three simulations (i.e., EXEMSI, SFINCSS, HUBES), which were longer simulations than CAPSUL. Eskov concludes that observed changes in efficiency require at least a two-to-three-week isolation period in order to become detectable. Second, instances of team conflict reported varied over time in the teams studied, but by 40% of the mission completion point (in our data set equivalent to at least 90 days of a mission), all teams had reported at least one conflict within the team or with MC. Third, the psychological closing phenomenon was observed for the "average commander mission report length to MC per week over time." This may suggest that the dynamics impinging on team functioning later in a mission are different for crews than those earlier in a mission. Finally, in regards to team mood, the crew of Mars-500 reported comparatively higher total mood disturbance than other simulation teams. Data reported on team mood from two LDSEM-analog studies that included teams for a year or more (e.g., Mars-500, and a 1-year Antarctic team) showed a spike in total mood disturbance at the one-year point. Taken together, it seems that team functioning is likely to vary over longer durations and that more insight into long-durations in ICE is necessary to close Team Gap 1. Thus, we recommend that data be collected in LDSEM-analog environments that are long duration. While the exact duration may depend on the specific phenomenon of interest, missions of upwards of 90 days or even 1 year are likely to provide unique insights to constructs such as conflict and team mood, respectively. Data from long-duration missions to lower-earth orbit such as the mission by astronaut Scott Kelly and cosmonaut Mikhail Kornienko could also provide invaluable data on team functioning over time.

TABLE 2 - *SUMMARY OF RESEARCH RECOMMENDATIONS - DATA COLLECTION IN LDSEM-ANALOG ENVIRONMENTS*

Conduct research in ICE environments specifically such as the Antarctic and chamber simulations (e.g., HERA)
Note. These are all valuable and of equal priority • Research on team functioning (e.g., team cohesion) specifically in ICE (e.g., Antarctic stations, HERA environment) is needed to better estimate the predictor and criterion relationships in this unique context relevant to LDSEM. • Quasi-experimental and experimental research in ICE is needed to help identify specific threats to team functioning and to help NASA better anticipate needed countermeasures during a team's life cycle. • Data should be collected in LDSEM-analog environments that are long duration. While the exact duration may depend on the specific phenomenon of interest, mission of upwards of 90 days or even 1 year are likely to provide unique insights into constructs such as conflict and team mood, respectively.

SPONSORING RESEARCH THAT ALLOWS FOR THE CULMINATION OF DATA

Although we were able to locate a fairly large number of studies, much of the data did not easily lend themselves to systematic quantitative aggregation. Analog research is primarily based on small sample sizes and often focuses on only one team; however, this is not surprising given the extreme circumstances that LDSEM-analogs represent. Given the logistical difficulties of collecting data in LDSEM-analog environments, it may be unrealistic to expect that a given study, particularly those most similar to the conditions expected for LDSEM (e.g., long duration, ICE) could generate enough data to calculate a between-team effect size. Thus, it is **essential** that research in LDSEM-analog environments is specifically conducted with integration across simulations in mind. We describe two paths that will allow for systematic integration across multiple studies: (a) collecting raw data in such a way that it can be analyzed and summarized across studies, (b) collecting data in such a way that an effect size can be generated for eventual comparison in a meta-analysis. We describe these in more detail in this section.

1. Collect data in LDSEM-analog environments that are directly comparable.

The majority of research in LDSEM-analog environments reviewed here examined only one team. These studies reflect invaluable knowledge from time-intensive chamber space simulations such as Mars-500. Because team effect sizes cannot typically be generated from these space simulation studies (the generation of standardized effects allow for comparability across studies), an alternative means of comparing the results across these studies is needed. Research conducted in space simulations, particularly those sponsored by other space agencies, has systematically utilized the same measures to collect data on team functioning. As an example, the Homeostat method has been used to collect data on team interaction and has systematically measured efficiency across multiple simulations (e.g., SFINCSS, Mars-105, HUBES). This continuity allowed us to provide figures reflecting the change in efficiency over time, as well as to determine upper and lower limits across more than one space simulation (e.g., Figures 2a and 2b). Assessing the comparability across studies in order to benchmark team functioning becomes significantly limited (e.g., comparison of z scores calculated within teams that mask relative changes between teams), when different measures or different response formats (e.g., some scales have a neutral, some do not) are used. As an example, despite the fact that data on team cohesion was available for 13 different analogs, a variety of team cohesion operationalizations were used, which made it difficult to directly compare data across simulations in a meaningful way.

We suggest two research priorities to ensure the comparability of data: (1) the use of a "common core" of measures that is collected across all NASA-funded analog studies, and (2) the use of a scale homogenization process in order to show equivalence across measures. We describe both of these priorities next.

First, a "common core", or a standardized set of measures of core team variables (e.g., team cohesion, team conflict), should be used across all NASA-sponsored analog environments. The figures and most of the findings presented in our report were only possible because of the continuity of the measures collected across multiple LDSEM-analogs by researchers such as Leon, Gushin, and Vinohkohdova. The majority of published chamber space simulation research to date that examines team functioning has been conducted at the IBMP of the Russian Academy of Sciences. We highly recommend that, at a minimum, the common core should include the measures of team functioning that have been and continue to be systematically collected in the ISS and in important simulations (e.g., MARS 500, SFINCSS) by team researchers at IBMP. By NASA incorporating a common core of measures that include those measures previously collected in other simulations, NASA and IBMP will be able to make direct comparisons of team functioning between more recent NASA-sponsored analogs, such as HERA, and the longer-duration chamber simulations that have been and will be conducted by IBMP (e.g., HUBES, SFINCSS, Mars-105, Mars-500). Because of the importance of researching teams in ICE for extended periods of time to understand the key threats, indicators, and life cycle for LDSEM crews, and the likelihood that team sample sizes for any given study conducted in such an environment will be small, we believe that the use of a common core will be the most important factor in determining the extent of our cumulative knowledge of team functioning for LDSEM in the years to come. The use of a common core that, at a minimum, includes the standardized team measures used by IBMP in the ISS and their simulations, is the most significant recommendation made in this report.

The use of a common core should not preclude the exploration of innovative ideas being researched and new measures being developed; however, the novel measures should be given in addition to a common core that assesses constructs such as team cohesion, team conflict and team performance, to name a few. We do not view the use of a common core and innovative research ideas to be incompatible because many of the measures collected as part of a common core in other simulations were collected once a week or once a month. Thus, in many circumstances, it should not be too taxing to collect both common core measures and innovative measures. Table 3 summarizes some potential measures to be administered as part of a common core; others could be added.

TABLE 3 - *SOME COMMONLY USED MEASURES IN LDSEM-ANALOG STUDIES*

Measure	Description	Constructs Measured	Example Analogs That Used the Measure
Homeostat Method to measure Csh, number of tasks solved, tactics of each crew member (e.g., attempts for leadership, follower tactics)	Crews solve a number of tasks with gradually increasing difficulty, determined by the degree of reciprocal influence of team members. The team has 180 seconds to solve a task, and the task can only be solved with the coordinated actions of the whole team.	Assesses group effectiveness (e.g., team efficiency) and role distribution	EXEMSI HUBES SFINCSS and others
Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988)	10-item self-report mood scales measuring independent dimensions of positive affectivity and negative affectivity. Adjectives are responded to on a 5-point scale.	Positive Affect (PA) Negative Affect (NA) Team-level mood is captured with the average PA or NA across the team.	Antarctic and Arctic Expeditions Hi-SEAS
Profile of Mood States (POMS; Schacham, 1983)	Captures individuals' mood on six dimensions using a 5-point self-report scale. To arrive at a total mood disturbance score, the first five subscales listed are summed and then the vigor-activity subscale is subtracted.	Total Mood Disturbance Tension-anxiety Depression-dejection Anger-hostility Fatigue-inertia Confusion-bewilderment Vigor-activity Team-level mood is captured with the average across the team.	Mars 500 SFINCSS Scott Base Palmer South Pole
Communication analysis of audio communication with MC, written reports to MC	Used to monitor relationship dynamics within the team and between the team and MC, emotional status of communicator, coping strategy of communicator (Gushin et al., 2012).	Average length of written report. Communication parameters listed in Gushin & Yusupova (2003), such as humor and primary demands for information. More recent (e.g., Mars-105) analyses include: needs, activity, social regulation, negation, etc.	Mars 105 EXEMSI HUBES ECOPSY SFINCSS ISS
Number of conflicts in a 30-day period	For MARS-500 (Basner et al., 2014) reported within-crew and between-crew and MC; through crew member self-report of whether there was a current conflict or conflict within the last 7 days.	Conflict	Scott Base (x3; Antarctic winter-overs) Mars 500

Personal Self-Perception and Attitudes (PSPA) (Gushin, Efimov, Smirnova, Vinokhodova, & Kanas, 1998)	Used to understand interpersonal relationships before, during, and after the mission.	Values Cohesion Self-perception Interpersonal perception Relationships in the group	Mars 105 Mars 500 ISS (Vinokhodova & Gushin, 2014)
Sociometric Questionnaire	Ratings by crew members of each crew member on interpersonal liking and other dimensions	Interpersonal liking Cohesion to name a few	SFINCSS Mars 105 Mars 500

Note. This list is not intended to be exhaustive. There may be additional measures that have consistently been administered across multiple LDSEM analogs. We did not examine the reliability or validity evidence of these measures.

We see the need for standardized measures for basic team constructs such as cohesion. Team cohesion was inconsistently operationalized and measured across simulations. A uniform measure of cohesion that could be used across simulations would be beneficial for making the data comparable across simulations. Salas, Grossman, Hughes, and Coultas (2015) provide some suggestions for a uniform measure of cohesion. Their review indicated that team cohesion had a stronger relationship with team performance when they included both social and task components and when analyses were performed at the team level. We imagine that collecting team cohesion data using various methods (e.g., sociometry, unobtrusive measures, self-report) is prudent as different measures may provide insights into different aspects of cohesion.

Sociometry has been used in several space simulations to measure cohesion, although the particular implementation of this approach has been inconsistent. For example, LeScanff, Bachelard, Cazes, Rosnet, and Rivolier (1997) allowed individuals to both choose and reject other team members in their ratings, whereas the work of Vinkohdova and colleagues (2001; 2012) only allows for positive choice. Further, there seems to be some distinction made by participants through sociometric data for task and relationship affinity. Specifically, ratings for the same team members on a friendship index (e.g., "who one would vacation") as compared to more work-focused sociometric ratings (e.g., "who would you want to participate with in a future experiment with simulated isolation") are different even when collected at the same time and when targeting the same individuals (see Vinokhodova et al., 2001). Data from SFINCSS indicated that sociometric ratings of whether an individual would want to be in a future simulation with the crew member was more correlated with the team efficiency then sociometric ratings of whether an individual would want to vacation with the other crew members. Sociometric data may also provide valuable insights into social integration, pairing, subgrouping, and isolation. A sociometric measure that includes the information collected by Vinokhodova and colleagues (2001, 2012) should be included in the common core.

Unobtrusive measures such as physical distance or time spent together (e.g., Kozlowski, Chang, & Biswas, 2014) could also be used, in addition to self-report measure of cohesion, and are currently being investigated in NASA-sponsored research. It is likely that these measures are capturing different elements of cohesion and may be good for different insights. For example, Likert scale rating of overall team cohesion may help researches gain quick insight into the team's sense of spirit (e.g., Roma, 2015); unobtrusive measures could provide early warning signs of subgrouping. How well unobtrusive measures related to the other measure of cohesion should be empirically tested, and efforts are currently underway in some NASA-sponsored research. Salas et al. (2015) provides insights into what needs to be done regarding the

development of unobtrusive measures of team cohesion, and work by Steve Kozlowski and colleagues is currently under way.

BHP has been partnering with a variety of researchers who are typically employed outside of NASA to research team functioning in LDSEM-analog environments. There are certainly benefits to that approach. At the same time, discontinuity of researchers can come with costs if NASA does not have access to complete data sets in a timely manner, particularly for variables measured as part of the common core. For example, if raw data were available across simulations, Bayesian analyses could be conducted to gauge the uncertainty associated with estimated predictor and outcome relationships; these results could promptly inform research priorities regarding risk reduction. Further, with Bayesian approaches and access to raw data, it would be possible to model the distributions of the data collected on small sample sizes and create better estimates than we could do here with limited data. These are just some of the possibilities for improved culmination of LDSEM-analog studies. Regardless of the specific approach used, it is essential that clear data sharing guidelines be established. We suggest a two-tiered data-sharing system that would allow BHP to summarize quantitative data on the common core as they see fit but would also allow time for researchers who are creating cutting-edge solutions and asking novel questions to publish their findings. Specifically, BHP should have access to raw data collected as part of the common core in all NASA-sponsored analog data collection efforts within a reasonable time frame, likely shorter than the typical journal publication timeframe. Specific innovations or new measures developed by individual researchers could be governed by a separate set of, and in some cases, more restricted data sharing guidelines than would govern the common core.

Second, given the difficulty in collecting LDSEM-analog research data, BHP may wish to make use of a scale homogenization process to transform data that may otherwise not be comparable in meaningful ways. For example, if team cohesion data is collected using one scale with a response format that includes a neutral option and another that omitted a neutral option, data cannot simply be rescaled and assumed to provide a meaningful comparison. Briefly, scale homogenization is a means for demonstrating equivalence in scale responses such that the values on one scale can be meaningfully compared to the values on another scale. This approach has been explored in the social science literature examining happiness (e.g., deJonge, Veenhoven, Kalmijn, & Arends, 2015). Specifically, these authors have used scale homogenization to make sense of data collected in different nations across time and with different scales and questions. deJonge et al. (2014) provide further detail on several methods of scale homogenization. These methods are used to systematically determine how response options should be transformed, for example, by examining the boundaries of the different response options (e.g., 1-1.8 on the first scale is equivalent to 1 to 3 on the second scale). For their purposes, deJonge et al. (2014) used the Reference Distribution Method, which involves collecting data on two different scales and then using the response distributions to inform the conversion of the data collected by different formats. Another approach is the Semantic Judgment of Work Value in Context Method, which involves SME judgment on the comparability of response options (see deJong et al., 2014 for a review).

Scale homogenization is relatively new and applications in the social sciences have been limited to self-report data. There could be novel applications of the scale homogenization process, which could enable integration of previous LDSEM-analog research. For example, data could be collected from a population with two different measures on the same construct and the distributions could inform a conversion algorithm that is used to transform results of data collected using different measures into a common data set. In addition, a scale homogenization

process could be used moving forward when researchers want to introduce novel measures to assess a common core construct. For example, if NASA-supported researchers wish to use a different measure to assess a variable collected by the common core (e.g., they suggest an improved measure, or prefer a different measure because it also assesses at the facet level of a variable), the researcher could be asked to demonstrate how responses to their measures equate to previously used measures in LDSEM-analog environments. This would facilitate linking team functioning data collected using new techniques with the team functioning data collected in previous simulations.

2. Collecting LDSEM-analog research so that data can eventually be meta-analyzed.

Meta-analysis can be used to effectively summarize research because data from a given study is represented by a common effect size metric (e.g., *r*, *d*). Effect sizes can be converted from one to another (e.g., *r* to *d* and *d* to *r*). We used r as our common effect size metric because the research was correlational (i.e., examining the strength and direction of a relationship rather than comparing mean differences across two or more conditions as is more commonly seen with experimental and quasi-experimental research). It is unlikely that all analog research can be designed so that it can contribute an effect size to a meta-analysis. For example, the single-team high fidelity simulations (e.g., Mars 500) provide invaluable insight into long-duration isolation and confinement, but they do not have a team sample size necessary for the calculation of an effect size so that can be included in team-level meta-analyses. However, in other cases such as HERA and Antarctic bases, research could be conducted in such a way that it can contribute a between-team effect size estimate. Where possible, analog research should be conducted in a way where it can contribute an effect size for later meta-analysis. This includes: (a) collecting data on enough teams to calculate a team-level effect size, and (b) having access to raw team-level data so that the normality assumptions can be tested, or alternatively, so that variance distributions can be specified, thus allowing for better-informed analyses (e.g., Bayesian analyses that can include an appropriate distribution for the variance parameters).

Our most common reason for not being able to generate a team-level effect size with the correlational data was the team-level sample size (i.e., team $n < 5$). We were not able to find guidance in the literature regarding a specific cutoff for the minimum sample size needed to generate an effect that can be meta-analyzed. Spearman's ρ can be calculated for as few as 3 teams; however, sampling variance must be calculated for the effect size to be incorporated into meta-analysis. Different meta-analytic approaches calculate sampling variance differently. Some approaches (e.g., Bayesian, Fisherian) calculate sample variance as $1/n$ and others as $1/(n - 3)$ (Brannick, 2001), and the Schmidt and Hunter (2015) method uses n-1 in the denominator of their random effects meta-analysis of correlations. As Brannick (2001) states, "if the sample is so small that the choice of n or n -3 is critical, then the researcher has a more serious issue to confront, namely, how to collect more data" (p. 469). Unfortunately for analog researchers, more data is not likely to be a feasible option for many studies. While differences in how sampling variance is calculated and the ability to calculate sampling variance at small sample sizes may generally be less of an issue in traditional meta-analyses, it is an important issue for the eventual culmination of team LDSEM-analog research. BHP may wish to explore the accuracy of the different meta-analytic approaches for use with extremely small sample sizes (e.g., correlations based on 3 to 7 teams) through simulation. It should be noted that while we used 5 as a cutoff for inclusion in our effect size data set, only a handful of studies in the current data set collected data on 3 or 4 teams. Thus, our decision to use a cutoff of 5 did not meaningfully affect our ability to broadly apply meta-analysis to the current data set. Moving forward, however, whether meta-analysis should be applied to correlations based on as few as 3

or 4 teams will be more important. The current HERA campaigns include 4 missions each which could possibly be incorporated into some types of meta-analysis (e.g., Hunter and Schmidt's psychometric meta-analysis) but not others.

Further, even if enough data are collected to allow an effect size to be generated, there is still a concern with assuming the variance is normally distributed (as we did in our weighted averages approach). As an example, a potential problem with small sample sizes is that the normality of the z-transformations, which are used in many meta-analytic approaches and in the weighted averages approach used in this study, is asymptotic. So, for very small sample sizes, the variance of z is not equal to $1/(n\text{-}3)$ and z is not normally distributed. We appreciate Alan Feiveson, who in addition to providing valuable comments on a previous draft, informally ran some simulations which suggested that normality and variance assumptions for the z-transformation of correlations were decent for $n = 7$, and even for $n = 5$ *if the original data were distributed as bivariate normal.* In other cases, the sample size has to be much larger for z-transformations to be approximately normally distributed with variance $1/(n\text{-}3)$. Access to raw data would allow for the normality assumption to be tested. Raw data would also allow for more sophisticated analyses than the weighted averages approach used here, in which the variance parameters that are consistent with the observed distribution of data are modeled.

Given the sample variance and normality issues noted, we make two recommendations. First, where possible, bivariate normality should be examined using raw data for the effect sizes provided by researchers to BHP. Alternatively, raw data could be provided to BHP (as discussed in the previous section) so BHP may examine the bivariate normality of the data before inclusion of a study in empirical Bayesian analyses or meta-analysis. Second, additional research should be conducted to better understand the implications of including effect sizes with a very small n in a meta-analysis.

Finally, there are other possibilities for generating effect sizes than correlations; one example is an effect size d calculated from repeated measures designs. Research on the efficacy of an intervention over time may be used to generate an effect size d. An effect size representing a between-team effect and an effect size representing a within-team effect of an intervention may not be conceptually similar, however, and the two may not be able to be combined into the same meta-analytic estimate. There are methods for accounting for differences in study design (e.g., repeated measures and independent group designs) on which effects are based (e.g., Morris & DeShon, 2002). The comparability of effect sizes and the types of analysis (e.g., meta-analysis) the effect sizes will contribute to should be explicitly planned prior to data collection.

TABLE 4 - *SUMMARY OF RESEARCH RECOMMENDATIONS - CONDUCTING RESEARCH THAT ALLOWS FOR QUANTITATIVE AGGREGATION*

Conduct LDSEM-analog research so that it better lends itself to quantitative aggregation
Given the logistical difficulties of collecting data in LDSEM-analog environments, it may be unrealistic to expect that a given study, particularly those most similar to the conditions expected for LDSEM (e.g., long duration, isolation and confinement), could generate enough data to calculate a between-team effect size. Thus, it is ***essential*** *that research in LDSEM-analog environments be specifically conducted with integration across simulations in mind.* **(A) Collect raw data in such a way that it can be analyzed and summarized across studies** Priority #1 Data on key constructs (e.g., team efficiency, communication, mood, and cohesion) should be collected with a common set of measures (i.e., a "common core") across all NASA-sponsored, LDSEM-analog environments (e.g., HERA). At a minimum, the team common core should include the team functioning measures that have been and continue to be systematically collected in important simulations (e.g., MARS 500, SFINCSS) conducted by IBMP, so that new data are directly comparable across simulations and with previously collected data. Priority #2 NASA should have access and rights to the common core data collected in NASA-sponsored LDSEM-analog environments within a reasonable time frame, likely shorter than the typical journal publication timeframe. Specific innovations developed by individual researchers could be governed by a separate set of, and in some cases more restricted, data sharing guidelines than would govern the common core. Priority #3 A scale homogenization process can be used to show equivalence across different measures used to operationalize the same construct, and to show the extent to which data collected with a new measure that reflects new developments can be compared to data collected with a previously used measure. **(B) Collect data in such a way that an effect size can be generated for eventual comparison in a meta-analysis.** Priority #1 When it is possible to collect data on multiple teams with the same or similar protocols so that a team-level effect size can be generated. Particularly with effects generated from small sample sizes, however, the bivariate normality of the original data should be examined. Even better than an effect size, however, is access to the raw data so that assumptions can be properly tested and appropriate distributions specified in more sophisticated analyses. Additionally, for research conducted with quasi-experimental or experimental designs, effect sizes could be based on repeated measures, but how the different effect sizes will be compared across studies should be considered ahead of time.

GAPS AND RESEARCH RECOMMENDATIONS REGARDING SPECIFIC TEAM FACTORS

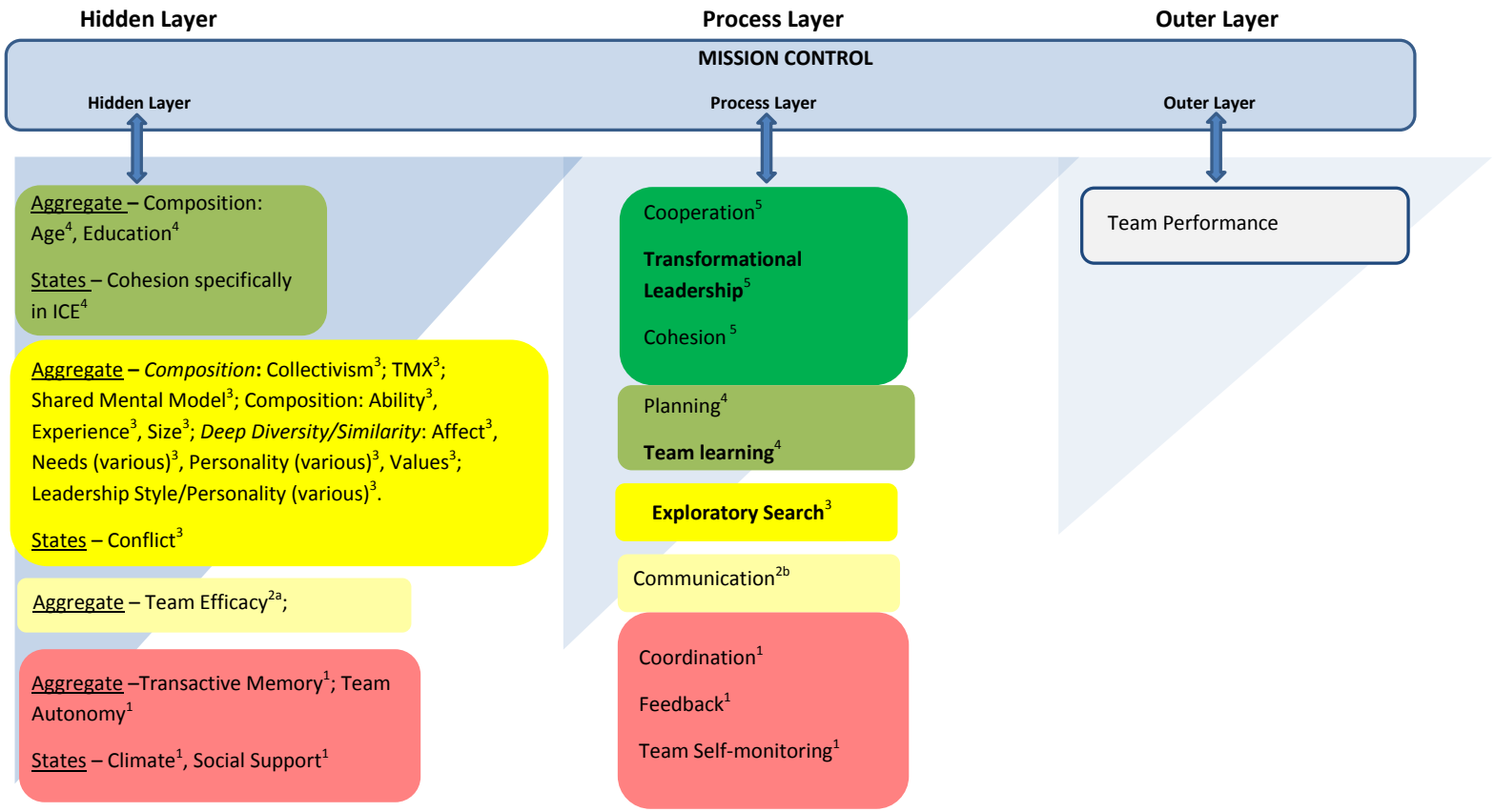
In the last section of the report, we provide some recommendations regarding specific areas in need of research by identifying gaps between the evidence we reviewed and a recently developed model of psychosocial factors in LDSEM (Schmidt, 2015). There are two caveats to the recommendations made in this section. First, although progress is being made, there is still a significant gap in our understanding of the key threats, indicators, and life cycles of team for LDSEM. Virtually all aspects of LDSEM team functioning are in need of research so that we can continue to accumulate knowledge across different teams with varying compositions and that are subjected to different conditions. Second, BHP has used the approach of conducting literature reviews and operational assessments (where relevant) of specific areas to better understand the research needed in those areas. Those reports can provide detail and develop the

theoretical rationale for specific hypotheses in need of research in the specific domain. As such, we keep our comments at a high level; however, these high-level comments should not be understood to mean that the specific priorities identified by previous and future reports are not important.

We summarize the type of LDSEM-analog evidence available for specific team factors in Figure 10, which is an adaptation of Figure 7 from Schmidt's (2015) model of psychosocial factors believed to affect team performance in LDSEM. In her Figure 7, Schmidt summarizes variables as an input layer (e.g., individual and situation context characteristics), a hidden layer (i.e., interactional and group factors), and a process layer (e.g., communication, coordination) believed to influence team performance. We adapted her model by including a link to MC. We also do not show the input layer in our figure for ease in viewing. We agree that the inputs she listed (e.g., personality, self-care, task type, familiarity) are important, however, our study examined these mostly at the team-level or as fidelity characteristics.

At the hidden layer, what Schmidt describes as aggregate and state-like factors have been explored within LDSEM-analog research. Our summary table reflects that within the hidden layer, we note that age and educational background composition and cohesion have been examined in LDSEM-analog research, but our empirical understanding of their effects on team performance is limited. In the case of team cohesion, this limited understanding is specific to ICE environments, as there seems to be a good understanding of the team cohesion and team performance relationship in non-ICE LDSEM-analogs (i.e., teams that live and work together for long durations such as firehouses and deployed military teams). In the figure, we note several other aggregate (e.g., leadership style) and state (e.g., conflict) factors have been explored within analog research but for which a relevant meta-analytic effect that could inform the prior distribution for the weighted averages could not be located. For example, we know the effect of task and relationship conflict on outcomes such as team innovation (Hülsheger, Anderson, & Salgado, 2009), but not on compatibility, as examined in analog research (Seymour, 1971). The weighted averages approach we used in this study could be applied to these effects as well. In order to do so, prior distributions need to be established; a number of methods could be used, including SME judgment, a meta-analysis of studies examining the specific effect in traditional teams, or another independent data source that could inform the prior distribution for these variables.

Regarding team processes, the size and direction of some effects seem well understood, as we were able to generate effect sizes for the relationship between team performance and variables such as cooperation. We note, however, that the local validity data were from LDSEM-analog teams that were not in isolated and confined environments (e.g., firehouses, deployed military teams). We see no immediate reason to expect a different relationship for cooperation and team performance in ICE settings as compared to the other LDSEM-analog settings; cooperation is likely to be important. It is likely, however, that the threats to cooperation in LDSEM will need to be examined specifically in ICE. Within this layer, our review identified three additional process variable-performance relationships that were investigated in non-ICE, LDSEM-analog environments, including the relationships between team performance and planning, transformational leadership, and exploratory search. Of these, the transformational leadership and team performance relationship was estimated to surpass a small effect with 95% certainty. Next steps may be to determine if this relationship generalizes specifically to ICE LDSEM-analog environments; if so, transformational leadership should be incorporated in Schmidt's (2015) model of psychosocial adaptation. Additional variables not included in Schmidt (2015) but identified in our review can be found in Appendix C.



1	Unable to locate analog studies that provide quantitative evidence about how the team factor influences processes or outcomes.
2	Variables included in Schmidt (2015) that were either studied with fewer than 5 teams[a]; or, that were not linked to a second team functioning variable[b] (hence benchmarked**).
3	Able to generate an effect size for this relationship in an analog environment but could not locate a meta-analytic study to inform the weighted averages analysis.
4	Able to generate an effect size for this relationship in an LDSE-analog environment and locate a meta-analytic study that could inform the weighted averages analysis . There seems to be little empirical understanding of the relationship as determined by imprecision in the posterior distribution generated.
5	Able to generate an effect size for the relationship in an LDSE-analog environment, locate a meta-analytic study that could inform the weighted averages analysis, and the estimated posterior distribution which suggests a solid understanding of the size and/or direction of relationship in the local population at 95% certainty. It should be noted that none of the LDSE-analogs that provided data in this category were in ICE, but they did live and work together for extended periods (e.g., firehouses, deployed military teams).

Figure 10. **Summary figure regarding team functioning evidence**

Figure Note. Figure adapted from Schmidt (2015) Figure 7.
* Variables in bold font indicate variables added based on our review.
**For a list of additional variables (i.e., not included in Schmidt [2015] but examined for effects over time within the extant literature), or for samples including fewer than 5 teams, please see Table 3 of Appendix C.

As summarized in Figure 10, there are a number of areas included in Schmidt's (2015) model that have not been studied in ICE environments. At the hidden layer, we were unable to identify or obtain data from ICE analog studies that examined the effects of transactive memory, team autonomy, and social support on performance. Additional research is needed to understand other hidden layer variables in order to achieve more precise estimates of their effects on team performance, and to allow meta-analytic summaries of the threats to (e.g., specific compositions), and life cycle of, team functioning. Some of these relationships are currently being investigated in NASA-sponsored research. At the process layer, research is needed on the effect of coordination, feedback, and team monitoring. While there are multiple studies looking at communication patterns over time, there is limited understanding of the effect of communication on team outcomes. This is particularly relevant given the atypical communication quality and amount expected between the crew and MC for long-distance missions. Other variables, such as cooperation, planning, and team learning, could also benefit from research specific to the ICE context.

In general, it seems that research should be prioritized when the nature of the relationship would be most likely to change as a function of the LDSEM environment. Many NASA-sponsored research efforts are currently underway (e.g., team cohesion, team composition) that can help inform Team Gap 1. Continuing these efforts is a high priority; our comments are intended to complement those efforts and expand the team domains covered by NASA-sponsored research. At a high level, we see affect management, conflict management, and the intersection between behavioral medicine and team function as particularly needed areas to inform our gap in understanding key threats, indicators, and the life cycle of a team for LDSEM. Some initial ideas are provided in the following paragraphs as well as our rationale for their importance.

Team affect is an umbrella term that includes several collective-level affective constructs, including bottom-up conceptualizations such as affective convergence and affective diversity, as well as top-down conceptualizations such as affective culture (e.g., expression and suppression norm) and affective dynamism (pattern of change that allows team affect to emerge and change; Barsade & Knight, 2015). The team science research supports relationship betweens team affect and cooperation, conflict, coordination, and performance (see Barsade & Knight, 2015 for a recent review). We were able to plot average team member mood over time with data from a number of LDSEM-analog teams. The data we examined indicated a spike in total mood disturbance around the one-year mark for both teams from which data were collected from for more than a year. Some teams indicated a third-quarter increase in negative affectivity.

Longer-durations in isolation and confined settings provide a unique circumstance in which affect and affect management may strongly influence team performance. Research on affect in LDSEM-analog environments exists; however, much of it is focused at the individual level (for a notable exception see Šolcová, Lacev, & Šolcová, 2014). Applying a team lens and conducting research on team affect and team affect management allows for a more complete understanding of the role of affect in LDSEM team performance and crew member well-being. For example, team affect tends to move toward homogeneity via mechanisms such as emotional contagion (Totterdal, Kellett, Teuchmann, & Briner, 1998). The isolation and confinement expected for LDSEM could magnify mechanisms like emotional contagion. Further, crew composition factors, such as national diversity, could influence the emergence of team mood as well as the effectiveness of affect management. For example, different cultural (or space agency) norms for affective suppression or sharing may influence the efficacy of affect management approaches. Finally, little is known about how specific mood experiences influence the development and life cycle of team mood over time. Recent research conducted on the Mars-500 crew explored some of these issues

(i.e., mood intensity, mood contagion, emotion regulation; Šolcová et al., 2014). Crew mood was observed to be generally stable, except for two specific episodes. Mood synchrony between crew members occurred later and mostly in connection with specific events in the simulation (e.g., post-landing on Mars); instead, crew members mostly balanced one another (when one crew member reporting more intensity, another reported less intensity). Mars-500 crew members used differing emotion regulation strategies. Finally, emotions reported by crew members were different during isolation compared with pre-isolation (Šolcová et al., 2014). Given the likelihood of different dynamics in ICE, further exploration of team affect, emotions, emotion regulation, and affect management in ICE across diverse crews similar to that expected for LDSEM is warranted.

It should be noted that most of the team mood data presented in this report were generated from aggregated individual-level data. Some ideas regarding bottom-up influences on team mood, such as the movement toward homogeneity or mood contagion, could likely be examined with existing data in order to provide some initial understanding of team mood emergence in LDSEM. The research on the coping mechanisms used by teams in LDSEM-analog environments (e.g., Leon & Sandal, 2003) could inform some initial ideas about effective approaches to social support and team affect management strategies. Thus, further exploration of existing data as well as new data collected in LDSEM-analog environments using a measure similar to the one used by Šolcová et al., (2014) could improve our understanding of the role of team affect in LDSEM.

A second area in need of research is conflict management. LDSEM provides a unique context for conflict management because of multiple avenues in which crewmembers may need to manage competing priorities or inconsistent priorities across the crew. For example, if multiple MCs are utilized, competing information may be provided to different crewmembers (e.g., exercise at this time, do not exercise at this time), which could lead to ambiguous priorities, and the diversity of crewmember backgrounds (e.g., professional, national background) could also lead to competing priorities (e.g., fix the space vehicle, complete the science experiment). The expected communication delays with MC, increased autonomy as a crew moves into deep space, and the anticipated mission length all likely increase the importance for the crew to effectively manage conflict.

We thought the decline in team efficiency during HUBES simulation, as well as the dip in team efficiency during one of the SFINCSS teams was interesting. Researchers (e.g., Sandal, 2001; Vinokhodova et al., 2001) suggested that the decline might have been a result of instability in or a lack of established leadership structure, and intra-team conflict. With a possible shift toward a different leadership structure (e.g., shared leadership) during LDSEM, identifying the conditions needed for teams to establish a workable leadership structure and means for ensuring crews diverse in culture and gender can effectively resolve status conflict (Bendersky & Hays, 2012) is an important avenue of research.

Beyond their effects on team performance, affect and conflict management are important because they will likely play a key role in a team's resilience. While we may not be able to anticipate each threat a specific LDSEM crew will experience, a LDSEM crew will, at some point, be met with challenges. A critical aspect of correctly composing, training, and providing countermeasure support to crews will include consideration of the crew's resilience. Team resilience is the capacity to withstand and recover from challenges, pressure, or stressors (Alliger, Cerasoli, Tannenbaum, & Vessey, 2015). Some of challenges the crew faces may be subtle changes that result in a less than ideal team state (e.g., the general decline in positive mood), while others are more acute (e.g., dispute related to the involvement of MC in conflict management). Regardless, team resilience will likely be critical to a crew's success. Future research should

measure crew resilience in response to specific manipulations of stressors as well as the resilience to the subtle changes that occur during a team's life cycle over long-durations.

TABLE 5 - *SUMMARY OF RESEARCH RECOMMENDATIONS - TEAM FACTORS IN NEED OF STUDY*

Gaps and research recommendations in regards to team factors in need of study
• Although progress is being made, there is still a significant gap in understanding the key threats, indicators, and the life cycle of a team for LDSEM. All areas are in need of continued research so that we can continue to accumulate knowledge across different teams with varying compositions operating under different conditions. Aspects of team functioning unique to the LDSEM context should be prioritized in ICE research. • Research on affect in LDSEM-analog environments exists; however, it is generally focused at the individual level. Applying a team lens to research on team affect (e.g., the extent to which and what factors amplify emotional contagion) and team affect management will allow for a more complete understanding of the role of affect in LDSEM team performance and crew member well-being. • We thought the decline in team efficiency during the HUBES simulation, as well as the dip in team efficiency during one of the SFINCSS teams was interesting. The instability or lack of an established leadership structure and conflict were noted by researchers (e.g., Sandal, 2001; Vinokhodova et al., 2001) as possible reasons for these performance effects. Given the possibility of shared leadership during LDSEM, an exploration of conditions necessary for teams to establish a workable leadership structure, triggers of status conflict, and approaches to resolving status conflict in multicultural crews is an important avenue of research. • The LDSEM life cycle suggests that at some point the crew will be met with LDSEM challenges. A critical component to composing, training, and providing countermeasure support to crews must include a consideration of team resilience. Future research should measure team resilience in response to specific manipulations of acute stressors (e.g., inclusion of MC to resolve a specific conflict) as well as in response to the subtle changes that occur during a team's life cycle.

REFERENCES

*References marked with one asterisk indicate effect size data set studies.
**References marked with two asterisks indicate benchmark data set studies.

Alliger, G. M., Cerasoli, C. P., Tannenbaum, S. I., & Vessey, W. B. (2015). Team resilience: How teams flourish under pressure. *Organizational Dynamics, 44,* 176-184.

**Allison, M. T., Duda, J. L., & Beuter, A. (1991). Group dynamics in the Himalayas. *International Review for the Sociology of Sport, 26,* 175-191.

Arthur, W., Jr., Bennett, W., Jr., & Huffcutt, A. I. (2001). *Conducting meta-analysis using SAS*. Mahwah, NJ: Erlbaum.

**Atlis, M. M., Leon, G. R., Sandal, G. M., & Infante, M. G. (2004). Decision processes and interactions during a two-woman traverse of Antarctica. *Environment and Behavior, 36,* 402-423.

Barsade, S. G., & Knight A. P. (2015). Group affect. *The Annual Review of Organizational Psychology and Organizational Behavior, 2,* 21-46.

Bartone, P. T., Johnsen, B. H., Eid, J., Brun, W., & Laberg, J. C. (2002). Factors influencing small-unit cohesion in Norwegian Navy officer cadets. *Military Psychology, 14,* 1-22.

**Basner, M., Dinges, D. F., Mollicone, D. J., Savelev, I., Ecker, A. J., Di Antonio, A., et al. (2014). Psychological and behavioral changes during confinement in a 520-day simulated interplanetary mission to Mars. *PloS One, 9,* 1-10.

Beal, D. J., Cohen, R. R., Burke, M. J., & McLendon, C. L. (2003). Cohesion and performance in groups: A meta-analytic clarification of construct relations. *Journal of Applied Psychology, 88,* 989-1004.

Bell, S. T. (2004). *Setting the stage for effective teams: A meta-analysis of team design variables and team effectiveness*. Unpublished doctoral dissertation, Texas A&M University, College Station.

Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *Journal of Applied Psychology, 92,* 595-615.

Bell, S. T., Brown, S. G., Abben, D. R., & Outland, N. B. (2015). *Team composition issues for future space exploration: A review and directions for future research. Aerospace Medicine and Human Performance, 86,* 548-556.

Bell, S. T., Brown, S. G., Outland, N. B., & Abben, D. R. (2015). *Critical team composition issues for long-distance and long-duration space exploration: A literature review, an operational assessment, and recommendations for practice and research* (NASA/TM-2015-218568). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. *Journal of Management, 37,* 709-743.

Bendersky, C., & Hays, N. A. (2012). Status conflict in groups. *Organization Science, 23,* 323-340.

**Bhargava, R., Mukerji, S., & Sachdeva, U. (2000). Psychological impact of the Antarctic winter on Indian expeditioners. *Environment and Behavior, 32,* 111-127.

**Binsted, K. (2015). Third-quarter syndrome and affect synchronization on long-duration space exploration analog missions. *Unpublished report obtained from the author.*

**Bishop, S. L., Kobrick, R., Battler, M., & Binsted, K. (2010). FMARS 2007: Stress and coping in an arctic Mars simulation. *Acta Astronautica, 66,* 1353-1367.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. West Sussex, United Kingdom: Wiley.

Brannick, M. T. (2001). Implications of empirical Bayes meta-analysis for test validation. *Journal of Applied Psychology, 86,* 468-480.

Burke, C. S., & Feitosa, J. (2015). *Team culture for long-duration exploration missions* (NASA/TM-2015-218587). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Burke, C. S., Stagl, K. C., Klein, C., Goodwin, G. F., Salas, E., & Halpin, S. M. (2006). What type of leadership behaviors are functional in teams? A meta-analysis. *The Leadership Quarterly, 17,* 288-307.

Curran, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short Form of the Profile of Mood States (POMS-SF): Psychometric information. *Psychological Assessment, 7,* 80-83.

**Décamps, G., & Rosnet, E. (2005). A longitudinal assessment of psychological adaptation during a winter-over in Antarctica. *Environment and Behavior, 37,* 418-435.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). The cognitive underpinnings of effective teamwork: A meta-analysis. *Journal of Applied Psychology, 95,* 32-53.

DeChurch, L. A., & Mesmer-Magnus, J. R. (2015). *Maintaining shared mental models over long-duration space exploration missions* (NASA/TM-2015-218590). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

deJonge, T., Veenhoven, R., Kalmijn, W., & Arends, L. (2015). Pooling time series based on slightly different questions about the same topic forty years of survey research on happiness and life satisfaction in The Netherlands. *Social Indicators Research,* 1-29.

Doll, R. E., & Gunderson, E. E. (1971). The influence of group size on perceived compatibility and achievement in an extreme environment. *Personnel Psychology, 24,* 305-310.

**Ehmann, B., Balazs, L., Fülöp, É., Hargitai, R., Kabai, P., Peley, B., et al. (2011). Narrative psychological content analysis as a tool for psychological status monitoring of crews in isolated, confined and extreme settings. *Acta Astronautica, 68,* 1560-1566.

*Emurian, H. H., & Brady, J. V. (1984). Experimental analysis of team performance effectiveness: Incentive and training factors. (Report submitted to the Office of Naval Research under Contract No. N00014-80-C-0467). Baltimore, MD: Johns Hopkins University School of Medicine.

**Emurian, H. H., Brady, J. V., Ray, R. L., Meyerhoff, J. L., & Mougey, E. H. (1984). Experimental analysis of team performance. *Naval Research Reviews, 36*, 3-19.

**Emurian, H.H., Emurian, C.S., & Brady, J.V. (1978). Effects of a pairing contingency on behavior in a three-person programmed environment. *Journal of the Experimental Analysis of Behavior, 29*, 319-329.

**Emurian, H.H., Emurian, C.S., & Brady, J.V. (1985). Positive and negative reinforcement effects on behavior in a three-person programmed environment. *Journal of the Experimental Analysis of Behavior, 44*, 157-174.

**Eskov, K. N. (2011). Evaluation of the efficiency of interpersonal interactions in a long-term isolation experiment. *Human Physiology*, *37*, 879-882.

Gibson, C., McIntosh, T., Mulhearn, T., Connelly, S., Day, E. A., Yammarino, F., & Mumford, M. D. (2015). *Leadership/Followership for long-duration exploration missions* (NASA/TM-2015-218567). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*, 3-8.

*Gunderson, E. K., & Nelson, P. D. (1963). *Measurement of group effectiveness in natural isolated groups*. (Report No. 63-16). San Diego, CA: Navy Medical Neuropsychiatric Research Unit.

*Gunderson, E. K., & Ryman, D. (1967). G*roup homogeneity, compatibility and accomplishment* (No. NMNRU-67-16). San Diego, CA: Navy Medical Neuropsychiatric Research Unit.

*Gunia, B. C., Sipos, M. L., LoPresti, M., & Adler, A. B. (2015). Sleep leadership in high-risk occupations: An investigation of soldiers on peacekeeping and combat missions. *Military Psychology, 27*, 197-211.

**Gushin, V. I., Efimov, V. A., & Smirnova, T. M. (1996). Work capability during isolation. *Advances in Space Biology and Medicine*, *5*, 297-307.

Gushin, V., Efimov, V. A., Smirnova, T. M., Vinokhodova, A. G., & Kanas, N. (1998). Subjects perceptions of the crew interaction dynamics under prolonged isolation. *Aviation Space and Environmental Medicine, 69,* 556-561.

**Gushin, V. I., Kolinitchenko, T. B., Efimov, V. A., & Davies, C. (1996). Psychological evaluation and support during EXEMSI, *Advances in Space Biology and Medicine, 5,* 283-295.

**Gushin, V., Shved, D., Vinokhodova, A., Vasylieva, G., Nitchiporuk, I., Ehmann, B., & Balazs, L. (2012). Some psychophysiological and behavioral aspects of adaptation to simulated autonomous Mission to Mars. *Acta Astronautica*, *70*, 52-57.

**Gushin, V., & Yusupova, A. (2003). Preliminary results of isolated crew's communication psychological analysis. In *Abstracts for 54th Congress of International Astronautical Federation* (Vol. 5), 1-7.

**Gushin, V. I., Zaprisa, N. S., Kolinitchenko, T. B., Efimov, V. A., Smirnova, T. M., Vinokhodova, A. G., & Kanas, N. (1997). Content analysis of the crew communication

with external communicants under prolonged isolation. *Aviation, Space, and Environmental Medicine*, *68*, 1093-1098.

**Gushin, V. I., Zaprisa, J. M., Smirnova, T. M., & Popova. (2001). Characteristics of Russian and non-Russian crew members' communication with external parties under prolonged isolation. In V.M. Baranov (Ed.), *Simulation of extended isolation: Advances and problems (pp. 85-100)*, Moscow, Russia: FirmSLOVO.

**Hagen, D. H. (1961). *Crew interaction during a thirty-day simulated spaceflight*: Preliminary study. School of Aerospace Medicine, Brooks AFB, TX.

**Hatch, H. G., Algranti, J. S., Mallick, D. L., Ream, H. E., & Stinnett, G. W. (1964). Crew performance during real-time lunar mission simulation (Tech Rep. No. NASA TN D-2447). Washington, DC: NASA.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.

Higgins, J. P., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539-1558.

Hirschfeld, R. R., & Bernerth, J. B. (2008). Mental efficacy and physical efficacy at the team level: Inputs and outcomes among newly formed action teams. *Journal of Applied Psychology*, *93*, 1429-1437.

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: A comprehensive meta-analysis spanning three decades of research. *Journal of Applied Psychology*, *94*, 1128-1145.

Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

**Johnson, J. C., Boster, J. S., & Palinkas, L. A. (2003). Social roles and the evolution of networks in extreme and isolated environments. *Journal of Mathematical Sociology*, *27*, 89-121.

**Kahn, P. M., & Leon, G. R. (2000). Group climate and individual functioning in an all-women Antarctic expedition team. *Journal of Human Performance in Extreme Environments*, *5*, 35-43.

**Kanas, N., Weiss, D. S., Marmar, C. R. (1996). Crewmember interactions during a Mir space station simulation. *Aviation, Space, and Environmental Medicine*, *67*, 969-975.

**Kass, R., & Kass, J. (2001). Team-work during long-term isolation: SFINCSS experiment group-006. In V.M. Baranov (Ed.), *Simulation of extended isolation: Advances and problems (pp. 124-147)*, Moscow, Russia: FirmSLOVO.

Kelly, J. R., & Barsade, S. G. (2001). Mood and emotions in small groups and work teams. *Organizational Behavior and Human Decision Processes*, *86*, 99-130.

Klein, K. J., Dansereau, F., & Hall, R. J. (1994). Levels issues in theory development, data collection, and analysis. *Academy of Management Review*, *19*, 195-229.

*Kniffin, K. M., Wansink, B., Devine, C. M., & Sobal, J. (2015). Eating together at the firehouse: How workplace commensality relates to the performance of firefighters. *Human Performance, 28,* 281-306.

*Knight, A. P. (2015). Mood at the midpoint: Affect and change in exploratory search over time in teams that face a deadline. *Organization Science*, *26*, 99-118.

*Ko, J. (2005). *Impact of leadership and team members' individualism-collectivism on team processes and outcomes: A leader-member exchange perspective* (Unpublished doctoral dissertation). The University of Arizona, Tucson, Arizona.

Kozlowski, S. W. J., & Bell, B. S. (2013). *Work groups and teams in organizations: Review update* [Electronic version]. Retrieved July 10, 2015, from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/articles/927

Kozlowski, S. W. J., Chang, C. H., & Biswas, S., & (2014, February). Capturing the dynamics of team work. Presented at the 2012 NASA Human Research Program Investigators' Workshop, Houston, TX

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3-90). San Francisco, CA: Jossey-Bass.

**Kraft, N. O., Inoue, N., Mizuno, K., Ohshima, H., Murai, T., & Sekiguchi, C. (2002). Psychological changes and group dynamics during confinement in an isolated environment. *Aviation, Space, and Environmental Medicine, 73,* 85-90.

**Krins, P. W. (2009). *Beyond the "right stuff": The role of group processes in isolated confined extreme environments* (Unpublished doctoral dissertation). The Australian National University, Australia.

**Leon, G. R., Atlis, M. M., Ones, D. S., & Magor, G. (2002). A 1-year, three-couple expedition as a crew analog for a Mars mission. *Environment and Behavior, 34,* 672-700.

**Leon, G. R., Kanfer, R., Hoffman, R. G., & Dupre, L. (1994). Group processes and task effectiveness in a Soviet-American expedition team. *Environment and Behavior*, *26,* 149-165.

**Leon, G. R., List, N., & Magor, G. (2004). Personal experiences and team effectiveness during a commemorative trek in the High Arctic. *Environment and Behavior, 36,* 386-401.

**Leon, G. R., Sandal, G. M., Fink, B. A., & Ciofani, P. (2011). Positive experiences and personal growth in a two-man North Pole expedition team. *Environment and Behavior*, *43*, 710-731.

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61,* 273-307.

**Le Scanff, C., Bachelard, C., Cazes, G., Rosnet, E., & Rivolier, J. (1997). Psychological study of a crew in long-term spaceflight simulation. *The International Journal of Aviation Psychology*, *7*, 293-309.

*Lim, B. C., & Klein, K. J. (2006). Team mental models and team performance: A field study of the effects of team mental model similarity and accuracy. *Journal of Organizational Behavior*, *27*, 403-418.

*Lim, B. C., & Ployhart, R. E. (2004). Transformational leadership: Relations to the five-factor model and team performance in typical and maximum contexts. *Journal of Applied Psychology*, *89*, 610-621.

**Luger, T. J., Stadler, A., Gorur, P., Terlevic, R., Neuner, J., Simonsen, O., et al. (2014). Medical preparedness, incidents, and group dynamics during the analog MARS2013 mission. *Astrobiology*, *14*, 438-450.

Mathieu, J. E., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of Management*, *34*, 410-476.

**Mizuno, K., Inoue, N., Kraft, N., Ohshima, H., & Sekiguchi, C. (2001). Relationship between aerobic work capacity and mood during SFINCSS-99 project. In V.M. Baranov (Ed.), *Simulation of extended isolation: Advances and problems* (pp. 370-379), Moscow, Russia: FirmSLOVO.

Morgeson, F. P. (2015). *Issues in maintain team motivation over long-duration exploration missions* (NASA/TM-2015-218586). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*, 105-125.

**Natani, K., Shurley, J. T., & Joern, A. T, (1973). Interpersonal relationships, job satisfaction, and subjective feelings of competence: Their influence upon adaptation to Antarctic isolation. In O. G. Edholm & E.K.E. Gunderson (Eds.), *Polar human biology* (pp. 384-400). London: Heinemann.

*Nelson, P. D. (1964). *Compatibility among work associates in isolated groups* (No. NMNRU-64-13). San Diego, CA: Navy Medical Neuropsychiatric Research Unit.

**Nicolas, M., & Gushin, V. (2014). Stress and recovery responses during a 105-day ground-based space simulation. *Stress and Health*. Advanced copy published online http://dx.doi.org/10.1002/smi.2565

**Nicolas, M., Sandal, G. M., Weiss, K., & Yusupova, A. (2013). Mars-105 study: Time-courses and relationships between coping, defense mechanisms, emotions and depression. *Journal of Environmental Psychology*, *35*, 52-58.

Newman, D. A., Jacobs, R. R., & Bartram, D. (2007). Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal of Applied Psychology*, *92*, 1394-1413.

Oser, R. L., McCallum, G. A., Salas, E., & Morgan Jr, B. B. (1989). *Toward a definition of teamwork: An analysis of critical team behaviors* (No. NTSC-89-004). Orlando, FL: Naval Training Systems Center.

Palinkas, L. A., Gunderson, E. K., & Burr, R. G. (1989). *Psychophysiological correlates of human adaptation in Antarctica* (No. NHRC-89-5). San Diego, CA: Naval Health Research Center.

Palinkas, L. A., Gunderson, E. K., Johnson, J. C., & Holland, A. W. (2000). Behavior and performance on long-duration spaceflights: Evidence from analog environments. *Aviation, Space, and Environmental Medicine, 71,* 29-36.

**Palinkas, L. A., & Houseal, M. (2000). Stages of change in mood and behavior during a winterover in Antarctica. *Environment and Behavior, 32,* 128-141.

Palinkas, L. A., Keeton, K. E., Shea, C., & Leveton, L. B. (2011). *Psychosocial characteristics of optimum performance in isolated and confined environments* (NASA Technical Report# NASA/TM-2011-216149).

**Paul, F. U., Mandal, M. K., & Ramachandran, K. (2013). Interpersonal feelings of personnel in polar environment. *International Journal of Research in Social Sciences, 3,* 265-277.

**Paul, F. J., Mandal, M. K., Ramachandran, K., & Panwar, M. R. (2010). Interpersonal behavior in an isolated and confined environment. *Environment and Behavior, 42,* 707-717.

**Roma, P. (2015). [Team Cohesion in Analog Environments]. Unpublished raw data.

**Rosnet, E., Cazes, G., & Vinokhodova, A. (1998). Study of the psychological adaptation of the crew during a 135 days space simulation. *Acta Astronautica, 42,* 265-272.

Rousseau, D. M. (1985). Issues of level in organizational research: Multi-level and cross-level perspectives. *Research in Organizational Behavior, 7,* 1-37.

**Sandal, G. M. (2001). Crew tension during a space station simulation. *Environment and Behavior, 33,* 134-150.

**Sandal, G. M. (2004). Culture and crew tension during an International Space Station simulation; Results from SFINCSS'99. *Aviation, Space, & Environmental Medicine, 75,* 44-51.

**Sandal, G. M., & Bye, H. H. (2015). Value diversity and crew relationships during a simulated spaceflight to Mars. *Acta Astronautica, 114,* 164-173.

**Sandal, G. M., Endresen, I. M., Vaernes, R., & Ursin, H. (1999). Personality and coping strategies during submarine missions. *Military Psychology, 11,* 381-404.

**Sandal, G. M., Vaernes, R., & Ursin, H. (1995). Interpersonal relations during simulated space missions. *Aviation, Space, and Environmental Medicine, 66,* 617-624.

Salas, E., Grossman, R., Hughes, A. M., & Coultas, C. W. (2015). Measuring team cohesion observations from the science. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 57,* 365-374.

Schmidt, L. L. (2015). *A model of psychosocial factors for long-duration spaceflight exploration missions* (NASA/TM-2015-218582). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Schmidt, F. L., & Hunter, J. E. (2004). *Methods of meta-analysis: correcting error and bias in research findings (2nd ed.).* Thousand Oaks, CA: Sage.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: correcting error and bias in research findings (3rd ed.).* Thousand Oaks, CA: Sage.

*Seymour, G. E. (1970). *The concurrent validity of unobtrusive measures of conflict in small isolated groups* (Report No. 71-16). San Diego, CA: Navy Medical Neuropsychiatric Research Unit.

Shacham, S. (1983). A shortened version of the Profile of Mood States. *Journal of Personality Assessment, 47,* 305-306.

Shea, C., Slack, K. J., & Keeton, K. E., Palinkas, L. A., Leveton, L. B. (2011). *Antarctica meta-analysis: Psychosocial factors related to long-duration isolation and confinement* (NASA/TM-2011-216148). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

Smith-Jentsch, K. A. (2015). *On shifting from autonomous to interdependent work: What we know and what we need to learn* (NASA/TM-2015-218563). Johnson Space Center, Houston, TX: National Aeronautics and Space Administration.

**Šolcová, I., Gushin, V., Vinokhodova, A., & Lukavský, J. (2013). Emotional energy, work self-efficacy, and perceived similarity during the Mars 520 study. *Aviation, Space, and Environmental Mmedicine, 84,* 1186-1190.

**Šolcová, I. P., Lačev, A., & Šolcová, I. (2014). Study of individual and group affective processes in the crew of a simulated mission to Mars: Positive affectivity as a valuable indicator of changes in the crew affectivity. *Acta Astronautica, 100,* 57-67.

**Steel, G. D. (2001). Polar moods third-quarter phenomena in the Antarctic. *Environment and Behavior, 33,* 126-133.

**Steel, G. D., & Suedfeld, P. (1991). Temporal patterns of affect in an isolated group. *Environment and Behavior, 23,* 749-765.

**Tafforin, C. (2013). The Mars-500 crew in daily life activities: An ethological study. *Acta Astronautica, 91,* 69-76.

**Tafforin, C., Vinokhodova, A., Chekalina, A., & Gushin, V. (2015). Correlation of etho-social and psycho-social data from "Mars-500" interplanetary simulation. *Acta Astronautica, 111,* 19-28.

Totterdell, P., Kellett, S., Teuchmann, K., & Briner, R. B. (1998). Evidence of mood linkage in work groups. *Journal of Personality and Social Psychology, 74,* 1504-1515.

Tyerman, A., & Spencer, C. (1983). A critical test of the Sherifs' robber's cave experiments: Intergroup competition and cooperation between groups of well-acquainted individuals. *Small Group Behavior, 14,* 515-531.

**Ushakov, I. B., Morukov, B. V., Bubeev, Y. A., Gushin, V.I., Vasil'eva, G. Y., Vinokhodova, A. G., & Shved, D. M. (2014). Main findings of psychophysiological studies in the Mars 500 experiment. *Herald of Russian Academy of Sciences, 84,* 106-114.

**van Baarsen, B. (2013). Person autonomy and voluntariness as important factors in motivation, decision making, and astronaut safety: First results from the Mars500 LODGEAD study. *Acta Astronautica*, *87*, 139-146.

Viechtbauer, W. (2005). Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics*, *30*, 261-293.

Viechtbauer, W. (2010). Conducting meta-analysis in R with the metafor package. *Journal of Statistical Software*, *36*, 1-48.

Viechtbauer, W., & Cheung, M. W. L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, *1*, 112-125.

**Vinokhodova, A. G., Bystritskaya, A. F., & Eskov, K. N. (2001). Influence of intergroup relations on group effectiveness and interpersonal relations under long-term isolation conditions. In V.M. Baranov (Ed.), *Simulation of extended isolation: Advances and problems (pp. 100-116)*, Moscow, Russia: FirmSLOVO.

**Vinokhodova, A. G., Bystritskaya, A. F., & Eskov, K. N. (2002). Inter-group relationship under simulated long-term isolation conditions. *Space Technology*, *22*, 103-111.

Vinokhodova, A. G., & Gushin, V. I. (2014). Study of values and interpersonal perception in cosmonauts on board the international space station. *Acta Astronautica*, *93*, 359-365.

**Vinokhodova, A. G., Gushin, V. I., Eskov, K. N., & Khananashvili, M. M. (2012). Psychological selection and optimization of interpersonal relationships in an experiment with 105-days isolation. *Human Physiology*, *38*, 677-682.

**Wang, Y., Jing, X., Lv, K., Wu, B., Bai, Y., Luo, Y., et al. (2014). During the long way to Mars: Effects of 520 days of confinement (Mars500) on the assessment of affective stimuli and stage alteration in mood and plasma hormone levels. *PloS One*, *9*, 1-9.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070.

**Wood, J., Lugg, D. J., Hysong, S. J., & Harm, D. L. (1999). Psychological changes in hundred-day remote Antarctic field groups. *Environment and Behavior*, *31*, 299-337.

APPENDIX A – LIST OF SEARCH TERMS

	Search Terms
Cluster 1	team OR teams OR crew OR crews OR group OR groups
Cluster 2	NASA OR ESA OR RATS OR NEEMO OR Antarctic OR Antarctica OR Arctic OR Polar OR ISS OR Hi-SEAS OR Mars 500 OR Mars OR desert OR cave OR caves OR Skylab OR ISRU OR “pavilion lake” OR “altitude chamber” OR McMurdo OR “south pole” OR summit OR astronaut OR “space exploration” OR FMARS OR “Arctic mars simulation” OR “Shuttle/Mir” OR “Mir space station” OR “Mars 105” OR “Amundsen-Scott” OR Moscow OR “Mars 520” OR Mariners OR submarines OR "special ops" OR "command and control" OR firefighters OR cadets OR bootcamps
Cluster 3	teamwork OR taskwork OR transition OR action OR interpersonal OR cooperation OR emergent states OR “transactive memory” OR “mental model” OR “shared cognition” OR learning OR cohesion OR “psychological safety” OR efficacy OR potency OR consensus OR autonomy OR integration OR analysis OR formulation OR planning OR specification OR monitoring OR coordination OR conflict OR motivation OR confidence OR emotions OR affect or mood OR “information sharing” OR communication OR interaction OR “workload sharing” OR mission OR strategy OR goal OR climate OR “shared mental model” OR performance OR effectiveness OR resilience OR viability

APPENDIX B – LIST OF SPECIFIC JOURNALS SEARCHED

Academic Emergency Medicine: Official Journal of the Society for Academic Emergency Medicine
Academic Medicine
Academy of Management Proceedings
Academy of Management Review
Acta Astronautica
Acta Psychologica
Administrative Science Quarterly
Advances in Space Research
Aerospace Medicine and Human Performance
Aerospace Research Central
African Journal of Business Management
Annual Review of Psychology
Applied Ergonomics
Aviation, Space, and Environmental Medicine
CA Magazine
Communications of the ACM
Computational Intelligence
Computers & Industrial Engineering
Current Directions in Psychological Science
Environment and Behavior
Ergonomics
European Journal of Operational Research
European Journal of Personality
European Journal of Work and Organizational Psychology
Group & Organization Management
Group & Organizational Studies
Group Dynamics: Theory, Research, and Practice
Herald of the Russian Academy of Sciences
Human-Computer Interaction
Human Factors
Human Resource Management
Human Resource Management Review
International Journal of Aviation Psychology
International Journal of Cross Cultural Management
International Journal of Human-Computer Studies
International Journal of Industrial Ergonomics
International Journal of Productivity Management and Assessment Technologies
International Journal of Selection and Assessment
Journal for Healthcare Quality: Promoting Excellence in Healthcare
Journal of Applied Psychology
Journal of Applied Social Psychology
Journal of Business and Psychology
Journal of General Psychology
Journal of Human Performance in Extreme Environments
Journal of Management Information Systems
Journal of Occupational and Organizational Psychology
Journal of Organizational Behavior
Journal of Personality and Social Psychology
Journal of Research in Personality
Journal of the American College of Surgeons

Journal of Work and Organizational Psychology
The Lancet
Management Science
McGill Journal of Medicine
The Mount Sinai Journal of Medicine
National Productivity Review (Wiley)
Occupational Psychology Review
Organization Science
Organization Studies
Organizational Behavior and Human Decision Processes
Organizational Dynamics
Performance Improvement Quarterly
Personnel Psychology
Personnel Review
Planetary and Space Science
Project Management Journal
Psychologist-Manager Journal
Quality & Safety in Health Care
Research in Organizational Behavior
Research in Personnel and Human Resources Management
Safety Science
Small Group Research
Space Technology
Team Effectiveness and Decision Making in Organizations
Theoretical Issues in Ergonomics Science
Translational Behavioral Medicine
World Journal of Surgery

APPENDIX C – LIST OF TEAM FACTORS & INCLUDED STUDIES

TABLE C.1. STUDY FIDELITY.

Study Setting	Author	Year	Crew Size	Reported Sample Demographics	Familiarity	Dangerous Environment	Mission Length (Days)
Meta-Analysis							
Natural - Antarctic Stations	Nelson	1964	Average: 4.2	Average age: 26; all men; Navy and weather and science personnel	Most members were familiar with one another	Yes	365
	Gunderson & Ryman	1967	Average: 18	Average age: 27; all men; variety scientist and construction maintenance	All members were unfamiliar with one another	Yes	365
	Gunderson & Nelson	1963	Range: 14-40	Average age: 27; all men; navy, officers, technicians, scientists	NR	Yes	365
	Seymour	1971	Range: 8-30	Average age: 26; all men; Scientist technicians military specialists	NR	Yes	240
Other - Combat Teams	Lim & Klein	2006	Average: 7.5	Average age: 19.1	NR	No	730
	Lim & Ployhart	2004	Average: 5	Average age: 19.3 sample; all men	NR	NR	NR
Other - Deployed Military Units	Gunia	2015	NR	All men; Army, 64% junior enlisted, 31% non-commissioned officers, 4% officers	NR	NR	270
Other - Firehouses	Kniffin, Wnsink, Devine	2015	NR	Mostly men; firefighters	NR	Yes	NR
	Knight	2015	Range: 10-17	Average age: 20.31: 86% men; all military cadets	NR	No	120
Special Operations Teams	Ko	2005	Average: 9.4	Average age: 22.8; all men	NR	NR	NR
Lab Teams	Emurian & Bradey	1984	Range: 2-3	Average age: 28.41; all men, except 1 group had a female	NR	No	10
Benchmark							
Lab	Emurian, Brady, Ray	1984	Average: 3	College background	NR	No	10
	Emurian, Emurian, Brady	1978	Average: 3	All men	All members were familiar with one another		10
	Emurian, Emurian, Brady	1985	Average: 3	Average age: 24; all men			10
	Gushin, Efimov, Smirnova	1996	Average: 4	3 males; 1 female			63
Natural-Amundson-Scott	Johnson, Boster, Palinkas	2003	Average: 28	19 men, 9 women; 100% civilians	All members were familiar with one another	Yes	255

Natural- Antarctic Expedition	Kahn & Leon	2000	Average: 4	Average age: 38.25; all American women; diverse professional background	NR	Yes	67
	Wood, Lugg, Hysong, Harm	1999	Average: 6	All men; 3 scientists, 3 engineers	Some member were familiar with one another	Yes	110
Natural - Antarctic Traverse	Atlis, Leon, Sandal	2004	Average: 2	Average age: 46; 2 women; teachers; mixed national background	All members were familiar with one another	Yes	97
Natural - Antarctic Winter over	Bhargava, Mukherjee, Sachdeva	2000	Average: 26	Average age: 36.44; all men	All members were familiar with one another	Yes	455
Natural - Antarctic Winter over Scott Base	Steel	2001	Average: 10	NR	NR	Yes	365
	Natani, Shurley, Joern	1973	Average: 21	Professions: Admin, technical, craftsman, & scientists; 62% Navy	All members were familiar with one another	Yes	240
Natural - Bering Bridge Expedition	Leon, Kanfer, Hoffman	1994	Average: 12	Average age: 36; 9 men, 3 women; Soviet and US	Some member were familiar with one another	Yes	61
Natural - Concordia	Krins	2009	Average: 10	2 women, 8 men; 4 Italians, 5 French	NR	Yes	240
Natural - Dumont Station	DeCamps & Rosnet	2005	Average: 27	Average age: 29.58	NR	NR	350
Natural - Expedition	Leon, Atlis, Ones, Magor	2002	Average: 6	Average age: 35; 3 men 3 women; diverse professional background; 3 Canadians, 2 Norwegian, New Zealander	Some member were familiar with one another	Yes	365
	Leon, List, Magor	2004	Average: 3	Average age: 34.67; 2 men 1 woman; diverse professional background; 1 Canadian, 2 Norwegian	All members were familiar with one another	Yes	46
	Leon, Sandal, Fink, Ciofani	2011	Average: 2	Average age: 33.5; all men; American ski racers	All members were familiar with one another	Yes	55
	Paul, Mondol, & Ramachandron (Panwar added for 2010)	2010, 2013	Average: 25	23 men; 1 woman; professional and technical background	All members were familiar with one another	Yes	420
Natural - High Arctic Weather Station	Steel & Suedfeld	1991	Average: 5	Average age: 34.2; 4 men, 1 woman; mixed professional background; 4 Canadians, 1 Argentinian; 1 captain in armed forces	All members were unfamiliar with one another	Yes	53
Natural - Himalayan Trek	Allison, Duda, Beuter	1991	Average: 9	All women; professionals scientist and non-scientists; 5 Americans, 2 Dutch, 1 British	Some member were familiar with one another	Yes	19

Natural - MDRS	Krins	2010	Average: 6	Average age: 32; all men	NR	No	14
Natural - Palmer	Palinkas & Hauseal	2000	Average: 8	Average age: 36.4; 6 men, 1 woman; Americans; 100% civilians	NR	Yes	365
Natural - South Pole	Palinkas & Hauseal	2000	Average: 21	Average age: 34.4; 48 men 15 women; Americans; 80.6% civilian	NR	Yes	365
Natural - Submarine	Sandal, Endresen, Vaernes	1999	Average: 20	Average age: 27.11; 1 woman	NR	Yes	40
Natural	Krins	2009	Average: 12.75	10 women; 15 men	NR	NR	26
Space Simulation - AF Sponsored Simulation	Hagen	1961	Average: 2	Average age: 33; 2 men; AF officers; American; 0% civilians	NR	No	30
Space Simulation - ECOPSY	Gushin	1997	Average: 3	Average age: 30; all men; Russians	NR	No	90
Space Simulation - EXEMSI	Eskov	2011	Average: 4	Average age: 29; 3 men, 1 woman	All members were familiar with one another	NR	60
	Gushin Kolinitchenko Efimov	1996					
	LeScanff, Bachelard, Cazes	1997	Average: 4	Average age: 29; 3 men, 1 woman; Scientists; Europeans	All members were familiar with one another	No	60
	Sandal, Vaernes, Ursin	1995	Average: 4	Average age: 29; 3 men, 1 woman			
Space Simulation - HERA	Roma	unpublished	Average: 4	Average age: approximately 30; mixed gender (roughly 50/50); Educational background: Bachelor's degree or higher; mixed professional background (medicine, engineering, science); all Americans.	Some member were familiar with one another (varies by team)	No	7 (2014) or 14 (2015)
Space Simulation - Hi-SEAS	Binsted	2015	Average: 6	4 women; 2 men; Russian, American, American European; 1 military	NR	No	120
	Binsted	2015	Average: 6	3 women; 3 men; Canadian American French; one military personnel			
Space Simulation - HUBES	Eskov	2011	Average: 3	Average age: 33.67; all men; 2 astronauts, 1 medical doctor; Russians	NR	No	135
	Gushin	1997					
	Rosnet, Cazes, Vinokhodova	1998					
	Sandal	2001					
Space simulation - ISEMSI	Sandal, Vaernes, Ursin	1995	Average: 6	Average age: 28; multi-national	All members were familiar with one another	No	28
Space simulation - Isolation Chamber in Japan	Kraft, Inoue, Mizuno	1987	Average: 8	6 males 2 females; students and a tv announcer; Japanese;	Some member were familiar with one another	No	7

Space simulation - Lunar Landing Simulation	Hatch, Algrant, Mallick	1965	Average: 4	Average age: 35.5; all men	NR	NR	42
Space Simulation - Mars 105	Gushin, Shved, Vinokhodova	2012	Average: 6	Average age: 32.7; all men; diverse professional background; 4 Russians, 1 German, 1 French		No	105
	Nicholas, Sandal, Weiss	2013					
	Nicolas & Gushin	2014					
	Vinokhodova, Gushin, Eskov	2012					
Space Simulation - Mars 2013	Luger, Stodler, Gorur	2014	Average: 14	Average age: 32.9; 10 men, 4 women; 11 Austrian, 3 Europeans	Most members were familiar with one another	Yes	30
Space Simulation - Mars 500	Basner, Dinges, Mollicone	2014	Average: 6	Average age: 32.16; all men; 1 commander, 2 physicians, 3 engineers; Russian, French, Italian, & Chinese	All members were familiar with one another	No	520
	Sandal & Bye	2015					
	Solcova, Gushin, Vinokhodova	2013					
	Solcova Lacev Solcova	2014					
	Tafforin	2013					
	Tafforin, Vinokhodova, Chekalina	2015					
	Ushakov, Morukov, Bubeev	2014					
	van Baarsen	2013					
	Wang, Jung, Lv	2014					
Space Simulation - MDRS	Ehmann, Balazs, Fulop	2011	Average: 6	Average age: 36.4; 5 males 1 female; Hungarian	NR	NR	13
Space Simulation - MIR Simulator	Kanas, Weiss, Marman	1996	Average: 3	Participants were in their 30s	All members were familiar with one another	No	135
Space simulation - NEEMO	Roma	unpublished	NR	NR	NR	NR	NR
Space simulation - PMARS	Bishop, Kobrick, Battler	2010	Average: 7	Average age: 29.14; 4 men 3 women; researchers; 3 Canadians 2 Americans 2 (No Suggestions)	NR	Yes	120

Space Simulation - SFINCSS	Eskov	2011	Average: 4	3 teams 2 with all men; 1 team with 1 woman; ages ranged from 27-48; medical doctors, mechanic; sport expert; Russians, German	NR	No	110 and 240
	Gushin, Zaprisa, Pustinnikova	2001					
	Kass & Kass	2001					
	Mizuno, Inoue, Kraft	2001					
	Sandal	2004					
	Vinokhodova, Bystritskaya, Eskov	2001					
	Vinokhodova, Bystritskaya, Eskov	2002					
Spaceflight - ISS	Gushin & Yusupova	2004	Average: 3	NR	NR	Yes	19
***Note**: NR = Data not reported in source. Only the first three authors are listed.*							

TABLE C.2. TEAM FACTORS AND EFFECT SIZES BY STUDY.

Study Setting	Authors	Year	Predictor	Criterion	Significant?	*n*	Effect Size (*r*)
Antarctic Stations	Nelson	1964	age homogeneity	compatability-sociometry	n	14	0.14
			age homogeneity	compatability-sociometry	y		0.68
			pay grade homogeneity	compatability-sociometry	n		0.27
			pay grade homogeneity	compatability-sociometry	n		0.47
			biographical background homogeneity	compatability-sociometry	n		-0.11
			biographical background homogeneity	compatability-sociometry	n		0.43
			interests homogeneity	compatability-sociometry	n		0.15
			interests homogeneity	compatability-sociometry	n		0.19
			need for achievement homogeneity	compatability-sociometry	n		0.27
			need for achievement homogeneity	compatability-sociometry	n		0.34
			need for prominence homogeneity	compatability-sociometry	n		-0.15
			need for prominence homogeneity	compatability-sociometry	n		-0.08
			compatability-sociometry	supervisor performance rating	n		-0.05
			compatability-sociometry	supervisor performance rating	n		-0.47
	Gunderson & Ryman	1967	compatibility that mixed conflict and climate	accomplishment	y	15	0.9
			age homogeneity	accomplishment	n		-0.014
			education homogeneity	accomplishment	n		0.132
			worship homogeneity	accomplishment	n		0.026
			urban-rural homogeneity	accomplishment	n		0.35
			number of hobbies homogeneity	accomplishment	y		0.61
			achievement personality homogeneity	accomplishment	n		0.24
			autonomy personality homogeneity	accomplishment	y		0.58
			nurturance personality homogeneity	accomplishment	n		0.31
			orderly personality homogeneity	accomplishment	n		0.28
			useful personality homogeneity	accomplishment	n		-0.07
			motivation personality homogeneity	accomplishment	n		0.31
			decisive personality homogeneity	accomplishment	n		0.12
			expressed inclusion values homogeneity	accomplishment	y		-0.45
			wanted inclusion values homogeneity	accomplishment	n		0.046
			expressed control values homogeneity	accomplishment	n		-0.418
			wanted control values homogeneity	accomplishment	n		0.182
			expressed affection values homogeneity	accomplishment	n		0.07
			wanted affection values homogeneity	accomplishment	n		-0.179
			friend description efficient homogeneity	accomplishment	y		0.551
			friend description sympathy homogeneity	accomplishment	n		0.321
			friend description caution homogeneity	accomplishment	n		0.094
			friend description optimism homogeneity	accomplishment	n		0.121

			age homogeneity	compatibility	n		-0.171
			education homogeneity	compatibility	n		0.171
			worship homogeneity	compatibility	n		-0.085
			urban-rural homogeneity	compatibility	y		0.471
			number of hobbies homogeneity	compatibility	y		0.761
			achievement personality homogeneity	compatibility	n		0.4
			autonomy personality homogeneity	compatibility	y		0.489
			nurturance personality homogeneity	compatibility	n		0.364
			orderly personality homogeneity	compatibility	n		0.275
			useful personality homogeneity	compatibility	n		0.046
			motivation personality homogeneity	compatibility	n		0.349
			decisive personality homogeneity	compatibility	n		0.257
			expressed inclusion values homogeneity	compatibility	n		-0.346
			wanted inclusion values homogeneity	compatibility	n		0.154
			expressed control values homogeneity	compatibility	n		-0.311
			wanted control values homogeneity	compatibility	n		0.182
			expressed affection values homogeneity	compatibility	n		0.232
			wanted affection values homogeneity	compatibility	n		0.004
			friend description efficient homogeneity	compatibility	y		0.524
			friend description sympathy homogeneity	compatibility	n		0.196
			friend description caution homogeneity	compatibility	n		0.098
			friend description optimism homogeneity	compatibility	n		0.229
	Gunderson & Nelson	1963	teamwork cooperation (time 1)	compatibility (time 1)	n	9	-0.14
			teamwork cooperation (time 2)	compatibility (time 2)	n		0.33
			teamwork cooperation (time 1)	efficiency (time 1)	y		0.76
			teamwork cooperation (time 2)	efficiency (time 2)	y		0.91
			teamwork cooperation (time 1)	achievement (time 1)	y		0.94
			teamwork cooperation (time 2)	achievement (time 2)	y		0.92
			teamwork cooperation (time 1)	egalitarian atmosphere (time 1)	n		-0.17
			teamwork cooperation (time 2)	egalitarian atmosphere (time 2)	n		0.19
			compatibility (time 1)	efficiency (time 1)	n		-0.57
			compatibility (time 2)	efficiency (time 2)	y		-0.68
			compatibility (time 1)	achievement (time 1)	n		-0.45
			compatibility (time 2)	achievement (time 2)	n		-0.66
			compatibility (time 1)	egalitarian atmosphere (time 1)	y		0.85
			compatibility (time 2)	egalitarian atmosphere (time 2)	n		0.34
	Seymour	1971	number of people per station	peer nomination positive chooser RS	y	20	-0.5
			number of people per station	peer nomination positive chosen RS	y		-0.48
			number of people per station	internal disagreement	y		-0.45
			number of people per station	per nomination positive chosen RS	y		-0.48
			number of people per station	compatibility	y		0.58

			conflict negative choosers	compatibility	y		-0.68
			conflict negative chosen	compatibility	y		-0.5
			conflict internal disagreements	compatibility	y		-0.75
Combat Teams	Lim & Klein	2006	taskwork mental models similarity	efficiency and quality combined	y	71	0.29
			teamwork mental models similarity	efficiency and quality combined	y		0.21
			taskwork mental models accuracy	efficiency and quality combined	y		0.42
			taskwork mental models accuracy	efficiency and quality combined	y		0.23
	Lim & Ployhart	2004	leader personality extroversion	training maximum performance efficiency and quality	n	39	0.11
			leader personality extroversion	training typical performance efficiency and quality	y		0.5
			leader personality consc	training maximum performance efficiency and quality	n		-0.05
			leader personality consc	training typical performance efficiency and quality	y		0.18
			leader personality neuro	training maximum performance efficiency and quality	n		-0.13
			leader personality neuro	training typical performance efficiency and quality	y		-0.56
			leader personality openness	training maximum performance efficiency and quality	y		-0.31
			leader personality openness	training typical performance efficiency and quality	y		0.37
			leader personality agreeableness	training maximum performance efficiency and quality	y		-0.21
			leader personality agreeableness	training typical performance efficiency and quality	y		0.28
			transformational leadership	training maximum performance efficiency and quality	y		0.6
			transformational leadership	training typical performance efficiency and quality	y		0.32
Deployed Military Units	Gunia	2015	general leadership	morale	n	46	0.24
			general leadership	unit cohesion	n		-0.161
			general leadership	sleep quantity	n		0.16
			sleep leadership	morale	y		0.57
			sleep leadership	unit cohesion	n		0.38
			sleep leadership	sleep quantity	y	44	0.413
Fire-houses	Kniffin, Wnsink, & Devine	2015	workload amount	cooperation	y	244	0.14
			eat together	cooperation	n		0.11
			cook together	cooperation	y		0.13
			cooperation	work group performance	y		0.43
			workload amount	work group performance	y		0.31
			eat together	work group performance	y		0.19
			cook together	work group performance	y		0.14
	Knight	2015	informal planning prior to start	team positive mood	n	33	0.25
			informal planning prior to start	team positive mood	n		0.42
			informal planning prior to start	team positive mood	n		0.34
			informal planning prior to start	team negative mood	y		-0.48

			informal planning prior to start	team negative mood	n		-0.43
			informal planning prior to start	team negative mood	n		-0.35
			informal planning prior to start	exploratory search	n		0.31
			informal planning prior to start	exploratory search	n		0.12
			informal planning prior to start	exploratory search	n		-0.06
			informal planning prior to start	time to completion and number of obstacles completed	n		0.25
			team experience	team positive mood	n		0.43
			team experience	team positive mood	n		0.42
			team experience	team positive mood	n		0.21
			team experience	team negative mood	y		-0.32
			team experience	team negative mood	n		-0.3
			team experience	team negative mood	n		-0.24
			team experience	exploratory search	n		0.23
			team experience	exploratory search	n		0.1
			team experience	exploratory search	n		-0.05
			team experience	time to completion and number of obstacles completed	y		0.45
			team ability	team positive mood	n		0.24
			team ability	team positive mood	n		0.2
			team ability	team positive mood	n		0.25
			team ability	team negative mood	y		-0.2
			team ability	team negative mood	n		-0.17
			team ability	team negative mood	n		-0.05
			team ability	exploratory search	n		-0.13
			team ability	exploratory search	n		-0.33
			team ability	exploratory search	n		-0.32
			team ability	time to completion and number of obstacles completed	y		0.22
			PANAS positive affect	team positive mood	n		0.38
			PANAS positive affect	team positive mood	n		0.22
			PANAS positive affect	team positive mood	n		0.22
			PANAS positive affect	team negative mood	n		-0.43
			PANAS positive affect	team negative mood	n		-0.19
			PANAS positive affect	team negative mood	n		-0.15
			PANAS positive affect	exploratory search	n		0.22
			PANAS positive affect	exploratory search	n		-0.14
			PANAS positive affect	exploratory search	n		-0.23
			PANAS positive affect	time to completion and number of obstacles completed	y		0.56
			PANAS negative affect	team positive mood	n		0.14

			PANAS negative affect	team positive mood	n		-0.31
			PANAS negative affect	team positive mood	n		-0.02
			PANAS negative affect	team negative mood	n		0.25
			PANAS negative affect	team negative mood	y		0.54
			PANAS negative affect	team negative mood	n		0.35
			PANAS negative affect	exploratory search	n		-0.26
			PANAS negative affect	exploratory search	n		0.14
			PANAS negative affect	exploratory search	n		0.32
			PANAS negative affect	time to completion and number of obstacles completed	y		-0.21
			team positive mood	team positive mood	y		0.48
			team positive mood	team positive mood	y		0.44
			team positive mood	team negative mood	y		-0.62
			team positive mood	team negative mood	n		-0.2
			team positive mood	team negative mood	n		-0.12
			team positive mood	exploratory search	n		0.32
			team positive mood	exploratory search	n		0.23
			team positive mood	exploratory search	n		-0.06
			team positive mood	time to completion and number of obstacles completed	y		0.28
			team positive mood	team positive mood	y		0.55
			team positive mood	team negative mood	y		-0.75
			team positive mood	team negative mood	y		-0.45
			team positive mood	exploratory search	n		0.06
			team positive mood	exploratory search	n		-0.29
			team positive mood	time to completion and number of obstacles completed	n		0.41
			team positive mood	team negative mood	y		-0.52
			team positive mood	exploratory search	n		0.05
			team positive mood	time to completion and number of obstacles completed	n		0.2
			team negative mood	team negative mood	y		0.62
			team negative mood	team negative mood	n		0.43
			team negative mood	exploratory search	n		-0.33
			team negative mood	exploratory search	n		0.08
			team negative mood	exploratory search	n		0.37
			team negative mood	time to completion and number of obstacles completed	n		-0.28
			team negative mood	team negative mood	y		0.49
			team negative mood	exploratory search	n		0.1
			team negative mood	exploratory search	n		0.34

			team negative mood	time to completion and number of obstacles completed	n		-0.4
			team negative mood	exploratory search	n		0.19
			team negative mood	time to completion and number of obstacles completed	n		-0.06
			exploratory search	exploratory search	y		0.66
			exploratory search	exploratory search	n		0.22
			exploratory search	time to completion and number of obstacles completed	n		0.22
			exploratory search	exploratory search	y		0.54
			exploratory search	time to completion and number of obstacles completed	n		-0.01
			exploratory search	time to completion and number of obstacles completed	n		0.02
Special Operations Teams	Ko	2005	collectivism	LMX	y	89	0.42
			collectivism	TMX	y		0.33
			collectivism	Cohesion	y		0.2
			collectivism	team performance general	y		0.23
			collectivism	viability	n		0.11
			collectivism	team performance general	n		0.14
			collectivism	team viability	n		0.02
			collectivism	team performance general	n		-0.05
			collectivism	team viability	n		0.13
			individualism	LMX	n		-0.18
			individualism	TMX	n		-0.15
			individualism	Cohesion	n		-0.19
			individualism	team performance general	n		-0.19
			individualism	viability	y		-0.24
			individualism	team performance general	n		-0.01
			individualism	team viability	n		0
			individualism	team performance general	n		-0.16
			individualism	team viability	n		-0.09
			transformational leadership	LMX	y		0.59
			transformational leadership	TMX	y		0.37
			transformational leadership	Cohesion	y		0.37
			transformational leadership	team performance general	y		0.54
			transformational leadership	viability	y		0.32
			transformational leadership	team performance general	n		0.17
			transformational leadership	team viability	n		0.09
			transformational leadership	team performance general	n		0.09
			transformational leadership	team viability	n		0.17

			LMX	TMX	y		0.48
			LMX	Cohesion	y		0.46
			LMX	team performance general	y		0.47
			LMX	viability	y		0.28
			LMX	team performance general	y		0.2
			LMX	team viability	n		0.11
			LMX	team performance general	n		0.02
			LMX	team viability	y		0.21
			TMX	Cohesion	y		0.78
			TMX	team performance general	y		0.69
			TMX	viability	y		0.58
			TMX	team performance general	y		0.25
			TMX	team viability	y		0.21
			TMX	team performance general	n		0.1
			TMX	team viability	y		0.2
			cohesion	team performance general	y		0.71
			cohesion	viability	y		0.6
			cohesion	team performance general	y		0.35
			cohesion	team viability	y		0.31
			cohesion	team performance general	n		0.12
			cohesion	team viability	y		0.31
Lab Teams	Emurian & Bradey	1984	time spent socializing	accuracy	NR	5	0.91
			time spent socializing	accuracy	NR	6	-0.61
			time spent socializing	accuracy	NR	7	-0.35
			time spent socializing	accuracy	NR		-0.6
			time spent socializing	accuracy	NR	6	-0.7
			time spent socializing	accuracy	NR	7	-0.48
			time spent socializing	accuracy	NR		-0.75
			time spent socializing	accuracy	NR		-0.39
			time spent socializing	accuracy	NR		-0.7

Note: *All studies used a correlational research design.*

TABLE C.3. TEAM FACTORS USED FOR BENCHMARKING BY STUDY.

Setting	Authors	Year	*n*	Variables
Lab	Emurian, Brady, Ray	1984	1	accuracy points
	Emurian, Emurian, Brady	1978	4	social activity voluntarily spending time together
				communication failure
	Emurian, Emurian, Brady	1985	4	social activity voluntarily spending time together
	Gushin, Efimov, Smirnova	1996	1	sum of different areas of performance
				psychological general work capability
				sub dimensions of performance
Natural - Amundson-Scott	Johnson, Boster, Palinkas	2003	3	global coherence v subgrouping
Natural - Antarctic Expedition	Kahn & Leon	2000	1	PANAS-negative affectivity
				PANAS-positive affectivity
				interpersonal stressors
				psychological stress like loneliness
				task stress like gear and food
				physical stress like frostbite
				coping relative frequency
				effectiveness
				cohesion
	Wood, Lugg, Hysong	1999	2	tension
				team emotional state
				team effectiveness
				individual-focused cognitive readiness (aggregated)
				moral (aggregated)
Natural - Antarctic Traverse	Atlis, Leon, Sandal	2004	1	team potency
				PANAS-negative affectivity
				PANAS - positive affect
				stress
				energy level
Natural - Antarctic Winter over	Bhargava, Mukerji, Sachdeva	2000	1	conflict - interpersonal sensitivity
				compliance acceptance of instructions
				life and work satisfaction
				cooperation-rapport
				alcohol intake
				smoking
				sleep disturbance

Natural - Antarctic Winter over at Scott Base	Steel	2001	2	pleasure
				arousal
				POMS - total mood disturbance
				POMS - vigor
				POMS - activity
	Natani, Shurley, Joern	1973	3	total instances of conflict
				number of radio messages from crew to MC
				mean fluctuations of interaction across professional subgroups
				fluctuation in intragroup and intergroup communication
Natural - Bering Bridge Expedition	Leon, Kanfer, Hoffman	1994	1	developing strong good feelings toward each other
				working together as a team
				helping each other cope
Natural - Concordia	Krins	2009	1	diversity culture
				shared identity
				friendship
				socialize
				distinctiveness
				advice
				social environmental flux
Natural - Dumont station	DeCamps & Rosnet	2005	1	total number of stress reactions
				Thymic stress
				occupational stress
				somatic stress reactions
				social stress
Natural - Expedition	Leon, Atlis, Ones,	2002	1	PANAS-negative affectivity
				PANAS-positive affect
				stress negative events reported
	Leon, List, Magor	2004	1	conflict and tension
				PANAS-negative affectivity
				PANAS-positive affectivity
				group efficiency
				satisfied with being a member of the group
				coping strategy
				significant events
	Leon, Sandal, Fink	2011	1	PANAS-negative affectivity
				PANAS-positive affectivity
		2010, 2013	1	feelings of affection directed toward others
				trust in competence

	Paul, Mondol, & Ramachandron (Panwar added for 2010)			significance
				other FIRO-B dimension of inclusion, affection, and control (2010)
Natural - High Arctic Weather Station	Steel & Suedfeld	1991	1	pleasure
				arousal
Natural - Himalayan Trek	Allison, Duda, & Beuter	1991	1	team cohesion - part of trek
				team cohesion - closely knit
				team cohesion - enjoy group
Natural - MDRS	Krins	2010	1	pro-diversity norm
			2	alienation general (reversed)
			1	identification
Natural - Palmer	Palinkas & Hauseal	2000	1	POMS - tension and anxiety
				POMS - depression
				POMS - anger
				POMS - fatigue
				POMS - confusion
				POMS - vigor
Natural - South Pole	Palinkas & Hauseal	2000	1	POMS - tension and anxiety
				POMS - depression
				POMS - anger
				POMS - fatigue
				POMS - confusion
				POMS - vigor
Natural - Submarine	Sandal, Endresen, Vaernes	1999	1	stress social factors
				stress leadership and workload
				stress homesickness
				stress feeling isolated
				stress subjective health complaints
Natural	Krins	2009	1	shared identity
Space Simulation - AF-Sponsored Simulation	Hagen	1961	1	within team communication amount
Space Simulation - ECOPSY	Gushin, Zaprisa, Kolinitchenk	1997	1	commander report to MC length
				number of negative statements

Space Simulation - EXEMSI	Eskov	2011	1	efficiency CSh homeostat
	Gushin, Kolinitchenko, Efimov	1996	1	length of commanders report
				content of commander report (e.g., jokes)
	LeScanff, Bachelard, Cazes	1997	1	choose and reject members
				sociometry - friendliness
				Matrix of intra and inter-processes in groups (MIPG)
	Sandal, Vaernes, Ursin	1995	1	tension and anxiety expressed relative to total communication BALES
				jokes expressed relative to total communication BALES
				PA expressed relative to total communication BALES
				personality
				SYMLOG
Space Simulation - HERA	Roma	unpublished	6	cohesion
Space Simulation - Hi-SEAS	Binsted	2015	2	PANAS positive affect
				PANAS negative affect
Space Simulation - HUBES	Eskov	2011	1	efficiency CSh homeostat
	Gushin, Zaprisa, Kolinitchenko	1997	1	total duration of communication with MC session in seconds
	Rosnet, Cazes, Vinokhodova	1998	1	*Matrix of intra and inter-processes in groups (MIPG)*
	Sandal	2001	1	SYMLOG friendliness social distance
				SYMLOG friendliness w MC
				speech acts between crew members
				frequency of emotional expression jokes
				frequency of emotional expression negative
				frequency of emotional expression positive
				peer ranking
Space Simulation - ISEMSI	Sandal, Vaernes, Ursin	1995	1	tension and anxiety expressed relative to total communication BALES
				negative emotions expressed relative to total communication BALES
				jokes expressed relative to total communication BALES
Space Simulation - Isolation Chamber in Japan	Kraft, Inoue, Mizuno	1987	1	interpersonal liking
				dimensions of perceptions of others and self in terms of personality-like features
Space Simulation - lunar landing simulation	Hatch, Algrant, Mallick	1965	1	velocity error comparisons - translunar insertion
				individual-level performance

Space Simulation - Mars 105	Gushin, Shved, Vinokhodova	2012	1	average report length in words for the month
				average number of needs satisfaction statement in MC report
				average number of activity statement in MC report
				average number of social regulation statements in MC report
				activity
				positive mood
	Nicholas, Sandal, Weiss	2013	1	positive mood
				coping
				depression
				negative affect defense mechanism
	Nicolas & Gushin	2014	1	REST-Q total stress
				REST-Q total recovery
				sub dimensions of stress and recovery
	Vinokhodova, Gushin, Eskov	2012	1	team cohesion sociometry
				sociometry
				value orientations, self perception, and interpersonal relationships within the crew (PSPA)
Space Simulation - Mars 2013	Luger, Stodler, Gorur	2014	1	team affect
				stress
				group interactions(teamwork)
				number of incidence
Space Simulation - Mars 500	Basner, Dinges, Mollicone	2014	1	number of conflict with crew members
				Total mood disturbance
				POMS depression
				POMS vigor
				POMS confusion
				POMS tension
				POMS anger
				POMS fatigue
				individual mood and emotions
	Sandal & Bye	2015	1	deviation from mean (all values)
				values
	Solcova, Gushin, Vinokhodova	2013	1	individual work efficacy
				team identification and integration
				emotional energy
				individual mood and emotions
	Solcova, Lacev, Solcova	2014	1	interpersonal actions
				total median mood intensity

				total median negative mood intensity
				total median positive mood intensity
	Tafforin	2013	1	interpersonal actions
	Tafforin, Vinokhodova, Chekalina	2015	1	within crew interpersonal communication verbal
				non-verbal during breakfast
	Ushakov, Morukov, Bubeev	2014	1	number of radio messages from crew to MC
				MC messages to crew
	van Baarsen	2013	1	freedom of choice
				interest in going to Mars
				acceptance of individual suffering
				personal challenge of Mars trip
				facial expressions
	Wang, Jung, Lv	2014	1	POMS - total
				values - tradition
				deviation from mean (all values)
				additional values
Space Simulation - MDRS	Ehmann, Balazs, Fulop	2011	1	team spirit
				emotion pattern
Space Simulation - Mir simulator	Kanas, Weiss, Marman	1996	1	total mood disturbance
Space Simulation - NEEMO	Roma	unpublished	1	cohesion
Space Simulation - PMARS	Bishop, Kobrick, Battler	2010	1	positive mood
				negative mood
				PSS stress
				subdimensions of stress (by gender)
				coping (by gender)
Space Simulation - SFINCSS	Eskov	2011	1	efficiency CSh homeostat
	Gushin, Zaprisa, Smirnova,	2001	3	total duration of audio communication between commander and MC
	Kass & Kass	2001	3	*pre and post FIRO-b*
				leadership styles
	Mizuno, Inoue, Kraft	2001	3	POMS - tension and anxiety
				POMS - depression
				POMS - anger
				POMS - fatigue
				POMS - confusion
				POMS - vigor
	Sandal	2004	2	SYMLOG friendliness
				SYMLOG friendliness w MC

				number of problems experienced
				friendliness between groups
	Vinokhodova, Bystritskaya, Eskov	2001	3	efficiency CSh homeostat
				average group work activity homeostat
	Vinokhodova, Bystritskaya, Eskov	2002	3	sociometry vacation together group integrity index involvement in communication
				sociometry vacation together group integrity index
				sociometry - future simulation
				average group work activity homeostat
				sociometry average across individuals
Spaceflight - ISS	Gushin & Yusupova	2004	2	total communication per day

Note. *Only the first three authors are listed. Variables in italics are not included in the data provided to BHP because of a lack of similarity across studies to justify aggregation, etc. but could be coded from the article and added to the database. The work by Emurian et al. was experimental or quasi-experimental. Benchmarking figures could have been generated with data from the Emurian articles; they were not because changes in the variables should be interpreted with a detailed understanding of the different conditions and manipulations in the study which, for the most part, were not particularly relevant to LDSEM.*

APPENDIX D – WHY THE CURRENT DATA SET DOES NOT LEND ITSELF TO TRADITIONAL FORMS OF META-ANALYSIS

Meta-analysis is preferred for summarizing quantitative data across studies and generating cumulative knowledge about a set of studies; the benefits over narrative reviews have been widely noted (See Glass, 1976; Arthur, Bennett & Huffcutt, 2001; Schmidt & Hunter, 2015). There are three commonly used frequentist approaches to meta-analysis (e.g., random effects, fixed effects, and mixed models). Although our effect size dataset included 253 effect sizes, which can be a promising number in a meta-analysis at the team-level, the effects were largely dependent because they were generated from only 11 different sources (e.g., journal articles, technical reports). Statistical dependence between effect sizes was particularly pronounced in our study because we coded any relationship related to team functioning. Studies included reported multiple predictors, examined relationships over time, and investigated multiple outcomes. For example, Gunderson and Ryman (1967) examined 22 different facets of homogeneity across two different outcomes for a total of 44 effect sizes from the study. Other studies included only a few predictors but over multiple time points. For example, Knight (2015) examined predictors, such as team formation activity and team positive mood, on a number of outcomes, such as team exploratory search, collected over multiple time points, allowing us glean 86 different effect sizes from this one source. This resulted in a data set with most relationships having too few independent effects (< 5) to be meta-analyzed using random effects approaches. The exception to this was the team cohesion and team performance relationship, which was reported across 6 different studies with sample sizes larger than 5.

A small number of studies for a given effect is problematic for random-effects models because random-effects models assume the dispersion between studies is real and not due to sampling error. When there are a small number of studies, the between-studies estimate is likely to be unstable and both the summary effect and the confidence interval may provide a false sense of assurance (Bornstein et al., 2009). Thus, with a small number of independent effects for any given relationship, our data were not appropriate for a random-effects model meta-analysis.

A meta-analysis can also be conducted using fixed-effect models; this approach can be conducted using as few as two effect sizes. Few studies (if any) in our data set could be argued to be functionally equivalent and to be studying the same specific effect, however. When studies are not functionally equivalent, a random-effects meta-analysis is more appropriate. Thus, we conclude that frequentist approaches to meta-analysis were not appropriate for the data set (i.e., studies were not homogeneous enough for a fixed-effects model and there were not enough studies on a particular relationship for a random-effects model).

Given this, we calculate a weighted average consistent with the empirical Bayesian estimation approach forwarded by Brannick (2001) and Newman et al. (2007), as a minimum-variance estimator, to understand our effect size data set and to benchmark team functioning in LDSEM-analog environments. We relied on this approach to use all available information to create the best possible estimate of local validity (e.g., the relationship between an examined predictor and criterion relationship in the LDSEM-analog environment) and to quantify our uncertainty with that estimate. Details of the empirical Bayesian estimation approach are provided in Brannick (2001) and Newman et al. (2007) for the interested reader.

APPENDIX E: ACRONYMS

BHP	Behavioral Health and Performance Element
Csh	efficiency metric included in the Homeostat measure
DRM	Design Reference Manual
EXEMSI	Experimental Campaign for the European Manned Space Infrastructure
HERA	Human Exploration Research Analog
HI-SEAS	Hawai'i Space Exploration Analog and Simulation
HUBES	Human Behaviour in Extended Spaceflights
IBMP	Institute of Medical and Biological Problems
ICE	Isolated and Confined Environments
ISEMSI	Isolation Study for the European Manned Space Infrastructure
ISS	International Space Station
LDSEM	Long Duration Space Exploration Mission
LSDA	Life Science Data Archive
MC	Mission Control
NASA	National Aeronautics and Space Administration
NEEMO	NASA Extreme Environment Mission Operations
POMS	Profile of Mood States
SFINCSS	Simulated Flight of International Crew on Space Station
SME	Subject Matter Expert

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave Blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	November 2016	Technical Memorandum

4. TITLE AND SUBTITLE
Data Mining Review of Team Benchmark Studies Related to Long-Duration Space Exploration Missions

5. FUNDING NUMBERS

6. AUTHOR(S)
Suzanne T. Bell, Ph.D., Shanique G. Brown, M.A., Tyree D. Mitchell, M. A.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Lyndon B. Johnson Space Center
Houston, Texas 77058

8. PERFORMING ORGANIZATION REPORT NUMBERS
S-1230

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
TM-2016-219280

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Unclassified/Unlimited
Available from the NASA Center for AeroSpace Information (CASI)
7115 Standard
Hanover, MD 21076-1320 Category: 53

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
As part of its integrated research plan, NASA seeks to manage the risk of performance and behavioral health decrements that result from inadequate cooperation, coordination, communication, and psychological adaptation within a team. As a part of this risk, NASA has identified a specific gap in understanding the key threats, indicators, and life cycle of teams that will embark on autonomous, long-duration and/or long-distance space exploration missions (LDSEM), Team Gap 1. An increasing number of studies have examined teams in LDSEM-analog environments (e.g., chamber simulations, Antarctic winter-overs) and can help inform this gap. We conducted a quantitative review of existing studies benchmarking team factors in LDSEM-analog environments to summarize the existing quantitative evidence on team functioning in such environments and to identify gaps to address in future research. This is the final report for our project (NASA contract NNJ15HK18P). We describe the methodology we used, the results of our efforts, and important research priorities related to Team Gap 1 moving forward.

14. SUBJECT TERMS	15. NUMBER OF PAGES	16. PRICE CODE
human behavior; human performance; psychological effects; psychology; teams	84	

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited