Automated Software Engineering (ASE) Group

PhD-Program Preparation for Successful Post-PhD Career

Tao Xie

School of Computer Science
Peking University

https://taoxiease.github.io/

taoxie@pku.edu.cn

Advice Portal: https://sites.google.com/site/asergrp/advice

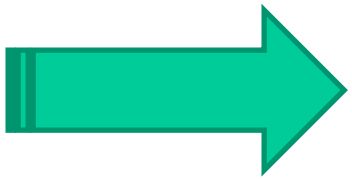
Based on **what** do I (you/advisor) consider myself (your student) ready to graduate with a PhD?

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Probl
    - Soluti
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

Synthesize a cohesive body of research to defend a thesis!

VDEC



high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Discussion Framework

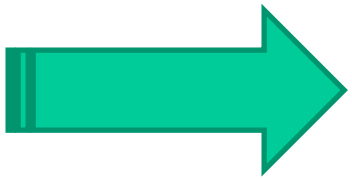
For each AVDEC skill

- **What** is the skill about?
- **How** to train/learn such skill?
  - E.g., practices conducted at ASE group
- What success **criteria** can be used to judge the accomplished level of the skill?
- How will such skill **impact** post-PhD career?

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC



high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Assessment – What

- Assessment of **others'** work
  - Later serving as program committee members or journal reviewers
  - Later advising more-junior collaborators in research development
- Assessment of **your own** work
  - Which ones are better among your *n* ideas?
  - What pros/cons of each of your *n* ideas?

Assessment – How

- Serve as **co-reviewer** of conference/workshop submissions reviewed by the advisor (as PC member)
  - Note that the advisor still needs to read and do the review himself/herself (likely based on the student's co-review)
- Review other students' **co-review**
  - Each submission has both primary co-reviewer and secondary co-reviewer
  - Secondary co-reviewer reviews primary co-reviewer's review & gives improvement suggestions

Assessment – How cont.

- After finishing your review, **compare** your review with the advisor's final review (and other students' reviews) to learn how to improve your future reviews

Assessment – Criteria

- **Closeness** of your review with your advisor's final review (e.g., complete, constructive, …)
- **Inclusion** of important points (in your review) that your advisor even does not think of, i.e., complementing your advisor's opinions
- …

Common patterns of students in initial phases

- Write very brief review comments (saying only good things)
- Write detailed comments on only writing issues
- Fall into two tendencies
  - not critical enough (accepting papers rejected by advisor)
  - over-critical (rejecting papers accepted by advisor)

Assessment – Impact

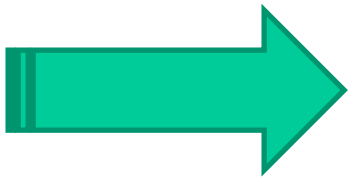
One of the two most critical among AVDEC along with **vision** (as faculty member)

- Advisor's role: **sounding board** (instead of idea generator?) for students
  - **Shaping** students instead of driving students
  - Analogy: search-based software engineering
    - **Fitness function**: advisor
    - Student's responsibilities: search strategies; production, mutation, crossover of chromosomes; …
- Serve on conference/workshop program committees or journal editorial boards/reviewers

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC



high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Vision – What

- If you are given 1 (4) **million dollars** to lead a team of 5 (10) team members for 5 (10) **years**, what would you invest them on?
- In 5 (10 or n) years, in what **X** (a research subarea), you want your name to be associated with?
  - Not just when mentioning your name, people think of X but when mentioning X, people think of your name!
- US NSF CAREER award topic is somewhat like that
- Not just a single paper but built/reflected with **N papers** along with **M talks/tutorials/conversations**

Vision – What cont.

Examples

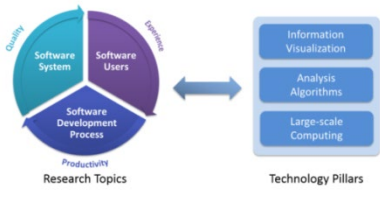
- Computational thinking (Jeannette Wing@MSR)
- Search-based soft eng (Mark Harman@UCL)
- Software analytics (Dongmei Zhang@MSRA/Tao)
- …

Vision – How

- Think and articulate underlying **assumptions** and **principles/themes** of your (past/ongoing) research - how do you (systematically) choose what to pursue?
  - core values that drive your research agenda in some broad way
- Figure out or form your **taste** of choosing problems (and solutions)

Example Principles

- Question **core assumptions** or **conventional wisdoms** about SE
  - D. Notkin: Software, Software Engineering and Software Engineering Research: Some Unconventional Thoughts. J. Comput. Sci. Technol. 2009
- Play/work around **industrial tools** to further improve them, e.g., Parasoft Jtest, MSR Pex/Pex4Fun/TouchDevelop
  - http://research.microsoft.com/en-us/projects/pex/community.aspx#publications
  - http://research.microsoft.com/en-us/projects/pex4fun/
  - http://research.microsoft.com/en-us/projects/touchdevelop/
- Collaborate with **industrial collaborators** to focus on high-practice-impact research problems and solutions, based on powerful infrastructure/data, e.g., MSR Pex/Pex4Fun/TouchDevelop, MSRA Software Analytics
  - http://research.microsoft.com/en-us/groups/sa/



Example Principles cont.

- Investigate SE mining **requirements** and adapt or develop mining algorithms to address them, e.g., Suresh Thummalapenta [ICSE 09, ASE 09]
  - http://researcher.watson.ibm.com/researcher/view_pubs.php?person=in-surthumm&t=1
- Cooperative testing/analysis: enabling effective and efficient cooperation of **tools** and **human** (for tasks traditionally tackled with tools/algorithms), e.g., Xusheng Xiao [Covana ICSE 11]
  - https://sites.google.com/site/xushengxiaoshome/
- **Integration** of static and dynamic analysis
- Using dynamic analysis to realize tasks originally realized by static analysis (or the other way around)
- …

Vision – How cont.

- Start writing your **Research Statement** (RS) earlier in your PhD program especially the **Future Work** section
  - Collect and learn from RSs released by those currently in the job market (from their homepages)
- Attempt to improve your answer to "What is your long-term research **vision**"?
- Be prepared to answer high-level questions such as "What is **beauty** of X"? (X being your chosen approach type such as dynamic analysis)

Vision – Criteria

- Satisfactory level of answering those **questions** earlier discussed in "What" slides
- Satisfactory level of "**Future Work**" in your Research Statement
- Make a successful submission to a **Visions and Challenges** track (such as the one for FSE 2014), different from the NIER track
- …

http://fse22.gatech.edu/cfp/visions-and-challenges

Vision – Impact

- A researcher with vision may provide more **inspiration/impact** to others

- A researcher with vision may become a **thinker** (thought leader), **strategist**, or **futurist** (beyond just a "doer")

Vision – Impact cont.

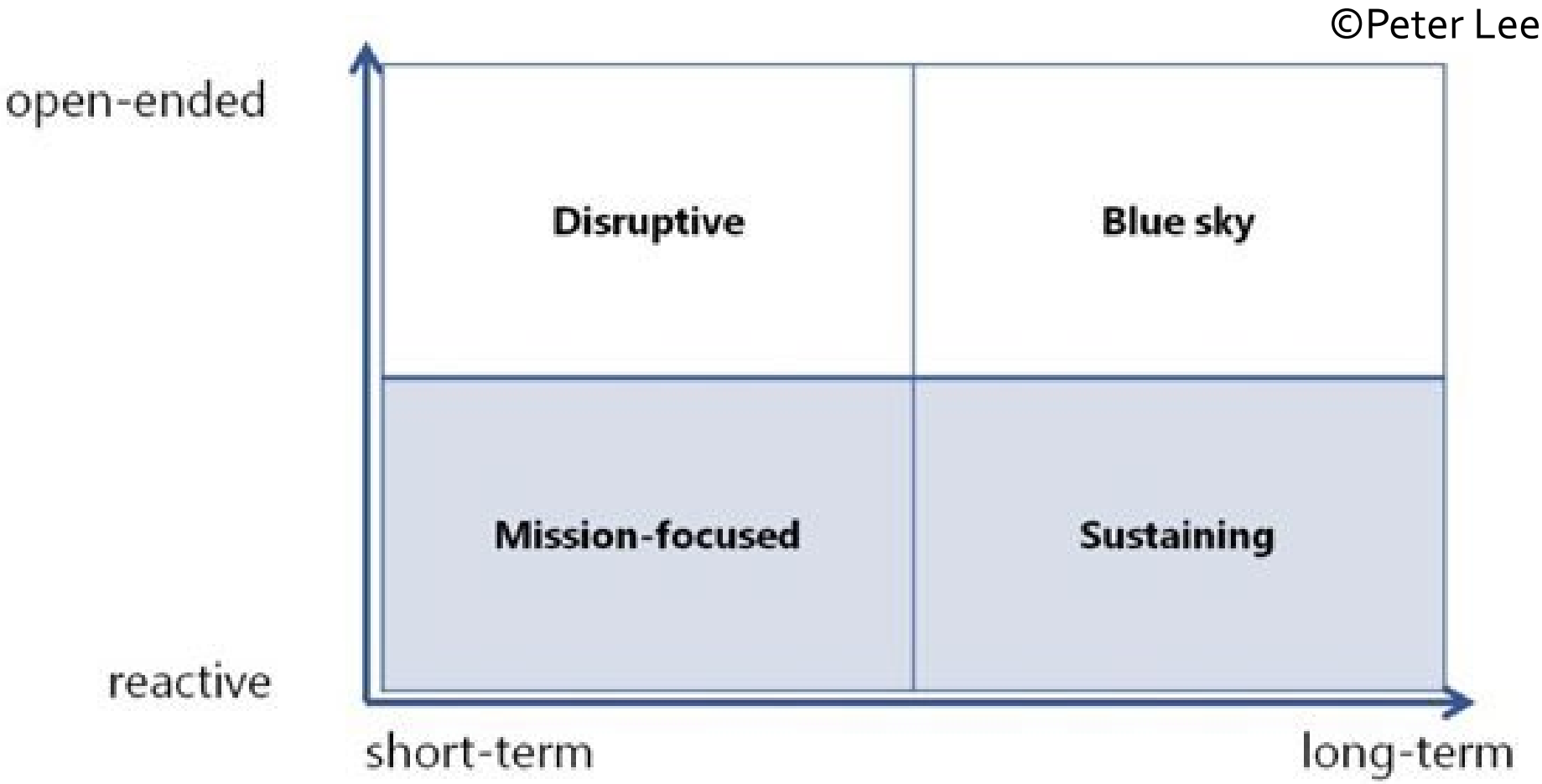
- Step back and think about what research problems will be most important and most influential/significant to solve in the long term
  - Long term could be the whole career
- People tend not to think about important/long term problems

	Less important	More important
Shorter term		
Longer term		√

Richard Hamming "**you and your research**"
http://www.cs.virginia.edu/~robins/YouAndYourResearch.html
Ivan Sutherland "**technology and courage**"
http://cseweb.ucsd.edu/~wgg/smli_ps-1.pdf

This slide was made based on discussion with David Notkin

Research Space



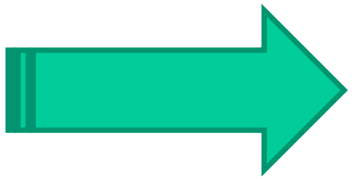
http://blogs.technet.com/b/inside_microsoft_research/archive/2011/12/31/microsoft-research-redmond-year-in-review.aspx a blog post by Peter Lee@MSR

Talk: The Pipeline from Computing Research to Surprising Inventions by Peter Lee@MSR
http://www.youtube.com/watch?v=_kpjw9ls14Q

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC



high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Design – What

- Pick what (good) problem to work on
  - Working on good problems or asking good questions is much appreciated in SE field

- Design what (good) solution/ideas to solve the problem

Compared with vision on ***n*** papers/projects, design is typically on ***one*** paper/project at a time

Design – How

Adopted in ASE group

- Document ideas in research idea log
  - You need to produce many ideas first before you accomplish producing new and good ideas
- Present in weekly group meeting for summarizing a specific sub-area
- Organize and participate small-group discussion
  - No advisor's participation
- Recommend/brainstorm papers to the advisor, who never read the papers
- Write down ideas in formal writing

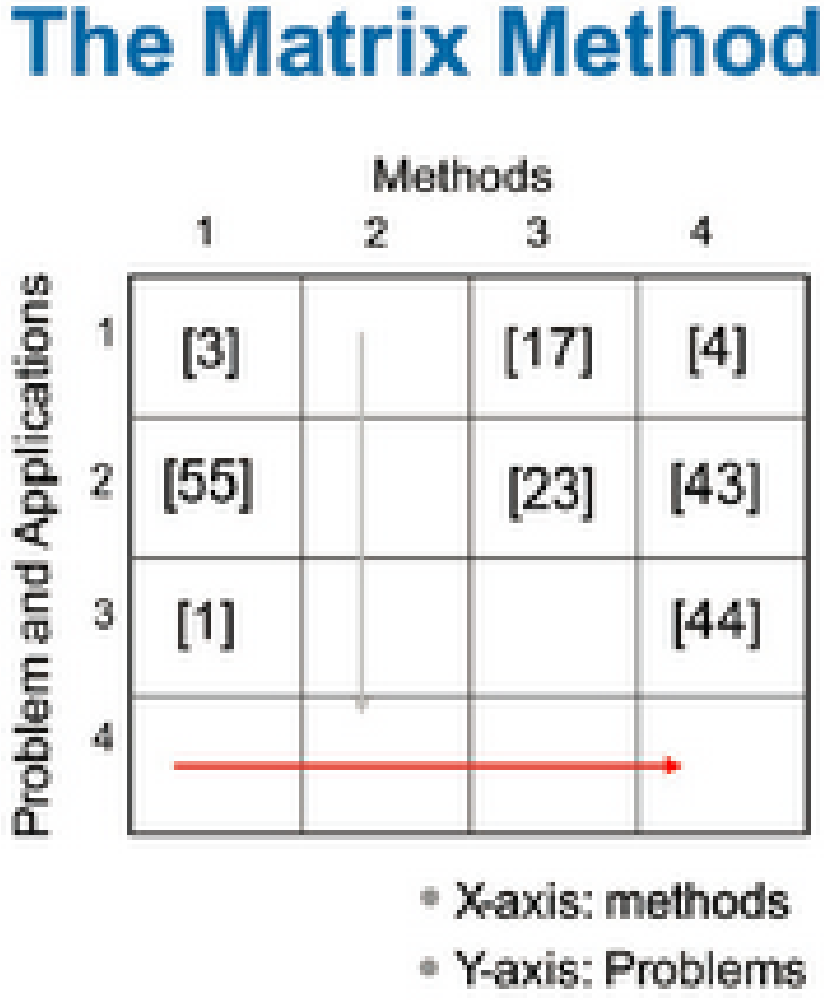
Get ***feedback*** from advisor on ideas

Design – How cont.

- Research Matrix (Charles Ling and Qiang Yang)
- Shallow/Deep Paper Categorization

More details, see my slides on "Mapping Out a Research Agenda" at https://sites.google.com/site/asergrp/advice

Technique: Research Matrix



See Book Chapter 4.3: Crafting Your Research Future: A Guide to Successful Master's and Ph.D. Degrees in Science & Engineering by Charles Ling and Qiang Yang

http://www.amazon.com/Crafting-Your-Research-Future-Engineering/dp/1608458105

Technique: Shallow Paper Categorization

- See ASE group's shallow paper category:
  - https://sites.google.com/site/asergrp/bibli
- Categorize papers on the research topic being focused
- Both the resulting category and the process of collecting and categorizing papers are valuable

Adopted by ASE group

Technique: Deep Paper Categorization

© Tao Xie

- Categorize papers on the research topic being focused (in a deep way)
- Draw a table (rows: papers; columns: characterization dimensions of papers)
- Compare and find gaps/correlations across papers

Adopted by ASE group

Example Table on Symbolic Analysis:

Paper	Category/Categories	Inputs	What to turn Symbolic	Manipulation of Symbolic Expressions/Path Conditions	Aggregation Across Runs	Outputs
DySy: Dynamic Symbolic Execution for Invariant Inference. Christoph Csallner, Nikolai Tillmann, and Yannis Smaragdakis. ICSE 2008	**Invariant Inference**	Existing Tests	Inputs from the existing test		Yes, since invariants stay the same after all transformations.	Inferred invariants from code. Abstract conditions over program variables that the concrete tests satisfy.
Precise Identification of Problems for Structural Test Generation. Xusheng Xiao, Tao Xie, Nikolai Tillmann, Jonathan Halleux. ICSE 2011.	**Problem Identification**	program under test	input parameters	User guidance???	Yes	

Design – Criteria

- For a high-quality paper published by you, did **you** pick the problem there?
  - Or **how much contribution** did you make in picking the problem?
- For a high-quality paper published by you, did **you** produce the ideas for the solution there?
  - Or **how much contribution** did you make in producing the ideas?
- For a successfully funded proposal, **how much contribution** did you make in picking/producing the problems/ideas there?
- …

Design – Impact

- All researchers need to pick good problems and produce good ideas to solve the problems
  - A big part of research innovations
- Faculty members need to have strong innovative capability in order to train students to be able to innovate, e.g.,
  - Collaboratively brainstorming new ideas with students
  - Implicitly guiding students to produce/own good ideas whose initial version was in the advisor's mind

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC

high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Execution – What

- Implement a tool **prototype** for the research
  - Tool development skill
  - Debugging/problem-solving skill
  - …
- Conduct **evaluation** to validate the research
  - Rigor skill
  - Analytic skill
  - …

Execution – What cont.

- **Team** management
  - Even if you are a single student working on your paper/project, your advisor (and other senior collaborators) is your team member managed by you
- **Risk** management
  - Identify, mitigate, manage risks
- **Task/time** management
  - Last-minute is not bad as long as you make the deadline and deliver high-quality deliverables
  - But often the time, you need help from your advisor (and collaborators) to improve your deliverable quality, and they are typically busy people

Execution – How

- Use **Gantt chart/Trello** as task management for your project and paper-deadline catching

- Write **weekly progress reports** to the advisor

Adopted by ASE group

Execution – Criteria

- How often do you **make deadlines** with high-quality submissions?
- How often do you need to **change** your **Gantt chart/Trello** along the way?
- How often do you **send** your drafts **late** to your advisor/senior collaborators, who couldn't spend time to give you feedback before submission?
- …

Execution – Impact

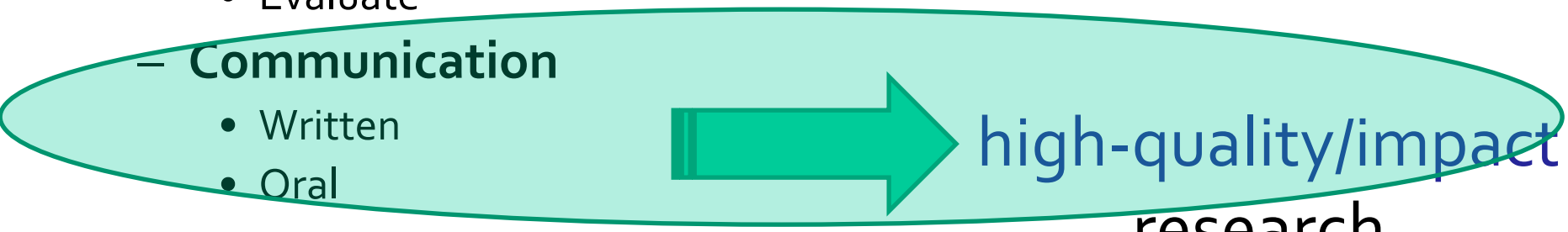
- All researchers need to carry out work effectively and efficiently
  - Be a reliable person to get things done!

- Faculty members especially often have many tasks at hand

Essential Skills for a (PhD) Researcher

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC



Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Communication – What

- Writing communication
  - Write a technical paper
  - Write a PhD thesis proposal (funding proposals later)
  - Write a PhD dissertation
- Oral communication
  - Present 5-min elevator talk
  - Present 25-min conference talk
  - Present 50-min seminar/job talk
- Listening comprehension
  - Understand what others say in Q/A of your talk, others' talks, causal conversations, …

(Written) Communication – How

- Write early and often
- Understand rationales of writing "rules" besides the "rules" themselves
  - See my slides on "technical writing issues", "how to write research papers" at https://sites.google.com/site/asergrp/advice
- Use writing defect logs
- Learn how to communicate ideas clearly (e.g., structured/top-down/outside-in and logical ways)
- Principle of advisor:
  - Never (re)write on students' papers except marking (on hardcopy); Iterate on giving guidance/explanation

Example Technique: Dotting Dragon Eyes

© Tao Xie

- Spot out "eyes" of a paper from ideas in the paper (e.g., general/interesting/inspiring ideas)
- Summarize the "eyes" in one phrase
- Promote the "eyes" in the paper title if possible
- Examples from our papers
  - **Context-Sensitive Delta Inference** for Identifying Workload-Dependent Performance Bottlenecks. Xusheng Xiao, et al. ISSTA 13
  - Automating Presentation Changes in Dynamic Web Applications via **Collaborative Hybrid Analysis**. Xiaoyin Wang, et al. FSE 12
  - **Iterative Mining** of Resource-Releasing Specifications. Qian Wu, et al. ASE 11
  - Alattin: Mining **Alternative Patterns** for Detecting Neglected Conditions. Suresh Thummalapenta, et al. ASE 09

Adopted by ASE group

Example Technique: "Balloon"/"Donut"

- Adopted by ASE group
- *Balloon*: the process is like blowing air into a balloon
- *Donut*: the final outcome is like a donut shape (with the actual realized problem/tool as the inner circle and the applicable generalized problem/solution boundary addressed by the approach as the outer circle)
- Process: do the following for the problem/solution space separately
  - Step 1. Describe what the exact concrete problem/solution that your tool addresses/implements (assuming it is X)
  - Step 2. Ask questions like "Why X? But not an expanded scope of X?"
  - Step 3. Expand/generalize the description by answering the questions (sometimes you need to shrink if overgeneralize)
  - Goto Step 1

Example Application of "Balloon"/"Donut"

© Tao Xie

- Final Product: Xusheng Xiao, Tao Xie, Nikolai Tillmann, and Jonathan de Halleux. Precise Identification of Problems for Structural Test Generation. *ICSE 2011*
- Problem Space
  - Step 1. (Inner circle) Address too many false-warning issues reported by Pex
  - Step 2. Why Pex? But not dynamic symbolic execution (DSE)?
  - Step 3. Hmmm… the ideas would work for the same problem faced by DSE too
  - Step 1. Address too many false-warning issues reported by DSE
  - Step 2. Why DSE? But not symbolic execution?
  - Step 3. Hmmm.. the ideas would work for the same problem faced by symbolic execution too
  - ….
  - Outer circle: Address too many false-warning issues reported by test-generation tools that focus on structural coverage and analyze code for test generation (some techniques work for random test generation too)

Example Application of "Balloon"/"Donut"

- Final Product: Xusheng Xiao, Tao Xie, Nikolai Tillmann, and Jonathan de Halleux. Precise Identification of Problems for Structural Test Generation. *ICSE 2011*
- Solution Space
  - Step 1. (Inner circle) Realize issue pruning based on symbolic analysis implemented with Pex
  - Step 2. Why Pex? But not dynamic symbolic execution (DSE)?
  - Step 3. Hmmm… the ideas can be realized with general DSE
  - Step 1. Realize issue pruning based on symbolic analysis implemented with DSE
  - Step 2. Why DSE? But not symbolic execution?
  - Step 3. Hmmm … the ideas can be realized with general symbolic execution
  - ….
  - Outer circle: Realize issue pruning based on dynamic data dependence (which can be realized with many different techniques!), potentially the approach can use static data dependence but with tradeoffs between dynamic and static

(Oral) Communication – How

When holding 1-1 meeting with the advisor

- Avoid using **slides**(??)
- Avoid using **examples** as first resort
- Avoid using **white board** as first resort

These three "avoid" are on the opposite of helping communication – the reason is for the advisor to debug the student's communication bug

  - Going from the failure to the bug location instead of temporary "healing" or "workaround"

(Oral) Communication – How

- **Recommend** papers to the advisor, who never read the papers before
  - Learn how to explain things clearly
  - Learn (through the advisor's questions) what are important to communicate when reading a paper or conveying your own paper
- **Present** your or other's work to the advsior and the group (in 1-on-1 or group meetings)
- **Practice**, practice, practice before a talk

Communication – Criteria

- How high percentage of writing was **written by you** within a high-quality paper?
- How high density of marks were **made by your advisor** within your high-quality paper?
- Satisfactory level of presenting 5/25/50 min **talks**
- Satisfactory level of understanding and answering **questions**
- **…**

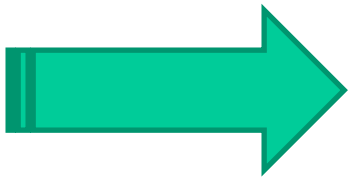
Communication – Impact

- Critical for faculty members
  - Write **proposals** (even when papers are primarily written by students)
  - Communicate research to **funding-agency managers** and **other researchers** in various meeting occasions
  - Advise **students** how to write better
- Critical for others too
  - Write **papers** (for researchers) and **tech docs**
  - Communicate with **team members** and **supervisors**
  - Communicate with **others** in various meeting occasions

Conclusion

- is able to independently
  - **Assessment**
    - Others' Work (e.g., conference PC members, journal reviewers)
    - Own Work
  - **Vision** (e.g., per n years, research agenda)
  - **Design** (e.g., per paper/project)
    - Problem
    - Solution
  - **Execution** (e.g., time/risk/team management)
    - Implement
    - Evaluate
  - **Communication**
    - Written
    - Oral

AVDEC



high-quality/impact research

Critical, Visionary, Creative, Strategic/Engineering, Logical… Skills

Thank you!

Questions ?

https://sites.google.com/site/asergrp/

Read more from my **advice portal**:
https://sites.google.com/site/asergrp/advice